

P

SEARCH REQUEST FORM
Scientific and Technical Information Center

Access DB# 54401

Requester's Full Name: _____ Examiner #: _____ Date: _____
An Unit: _____ Phone Number 30 _____ Serial Number: _____
Mail Box and Bldg/Room Location: _____ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): _____

Earliest Priority Filing Date: _____

For Sequence Searches Only Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.

STAFF USE ONLY

Searcher: _____
Searcher Phone #: _____
Searcher Location: _____
Date Searcher Provided: 11/7/01
Date Completed: 11/8/01
Searcher Prep & Review Time: _____
Client Prep Time: _____
On file Time: _____

Type of Search

NA Sequence (#) 1
AA Sequence (#) 1
Structure (#) _____
Bibliographic _____
Litigation _____
Fulltext _____
Patent Family _____
Other _____

Vendors and cost where applicable

STN _____
Dialog Reverse to NA
Questel Orbit _____
Or Link _____
Lexis Nexis _____
Sequence Systems 05
WWW Internet _____
Other (specify) _____

From: Ibrahim, Medina A.
Sent: Tuesday, November 06, 2001 1:33 PM
To: STIC-Biotech/ChemLib
Subject: 09/538, 396

please do the following search:

1. standard and interference search of a nucleotide sequence encoding SEQ ID NO:2 against all DNA databases.
2. oligo search of at least 30 contiguous bases of SEQ ID NO:1. Thanks

Medina A. Ibrahim
Patent Examiner
GAU:1638
CM1-9A12
Mailbox:9E12
(703)306-5822

Edward G. ...
Technical ...
STIC-Biotech
CM1-9A12

Pending Nucleic Acid and/or Pending Amino Acid database searches now generate two sets of results. These databases were split into two parts to reduce the time needed to update the databases daily. The split freed up more machine time for processing searches.

Searches run against the Nucleic Acid Pending database produce two sets of results, with the extensions, **.rnpm** and **.rnpn**

Searches run against the Amino Acid Pending database produce two sets of results, with the extensions, **.rapm** and **.rapn**

The Pending database search results should not be left in the case because they contain data that is confidential.

According to the Pre Publication Rules, every patent application received by the United States Patent and Trademark Office after November 29, 2000 will be pre-published at eighteen months from the effective filing date. When the application is published the contents, including the sequences, will become prior art.

Two new databases have been created to hold the pre-published sequences:
Published_Applications_NA contains nucleic acid sequences; the search results will have the extension .rnpb.

Published_Applications_AA contains amino acid sequences; the search results will have the extension .rappb.

Each pre-published application is given a unique Publication Number. An example of a Publication Number is US20021234567A1. The "US" indicates the application was a U.S. application. The first 4 digits show the calendar year the application was published. The next 7 digits represent when the application was published. This 7-digit number starts at zero at the beginning of each calendar year. Each application published is given the next number in order. The "A" indicates a utility patent application and the "1" shows that this was the first time the application had been published. If the applicant's submit changes to the application, they may request that the changed application be published again. In such instances, the "1" at the end of the number would be replaced by a "2".

Sequences in the PGPub database are public information; it is permissible to leave these results in the case.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2001, 21:39:51 ; Search time 148.21 Seconds
(without alignments)
6864.174 Million cell updates/sec

Title: us-09-538-396-1

Perfect score: 4492
Sequence: 1 aatcgagcagagtgatcc.....atctagcatcaaaaaaaaa 4492

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 351203 seqs, 113238999 residues

Word size : 30

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA:*

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
- 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
- 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
- 5: /cgn2_6/ptodata/2/ina/PTUS_COMB.seq:*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	ID	Description
------------	-------------	----	-------------

No matches found

Search completed: November 8, 2001, 00:54:46
Job time: 11695 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2001, 23:14:51 ; Search time 85.56 Seconds
(without alignments)
3456.429 Million cell updates/sec

Title: US-09-538-396-1

Perfect score: 4492
Sequence: 1 aatcgcgcacgcagtgatcc.....atctagcatalcaaaaaaaaaa 4492

Scoring table: OLIGO_MNC
Gapop 60.0 , Gapext 60.0

Searched: 84510 seqs, 32917641 residues

Word size : 30

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending_Patents_NA_New:*

- 1: /cgn2_6/ptodata/1/pna/PCT_NEW_COMB.seq:*
- 2: /cgn2_6/ptodata/1/pna/US06_NEW_COMB.seq:*
- 3: /cgn2_6/ptodata/1/pna/US07_NEW_COMB.seq:*
- 4: /cgn2_6/ptodata/1/pna/US08_NEW_COMB.seq:*
- 5: /cgn2_6/ptodata/1/pna/US09_NEW_COMB.seq:*
- 6: /cgn2_6/ptodata/1/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	ID	Description

No matches found

Search completed: November 8, 2001, 02:41:12
Job time: 12381 sec

[illegible]

```

17  oAspAsnLysAsnValIleThrPhePheLysProLeuThrLeuIleValG 34
   |||
342 GGACAAATAAAGACGTCACTACCTTCTTCAAGCCGCTACCCCTCATGCTTG 391
   |||
34  LProAsnGlyAlaGlyLysThrThrIleIleGlyCysLeuLysLeuSer 50
   |||
392 GCCCAAGGCGGTGGGCAAGCAGCATCATCGAGTGTGCTGAAGCTTCT 441
   |||
51  CysThrGlyGluLeuProProAsnSerArgSerGlyHisThrPheValH 67
   |||
442 TGCACCGCGCGCTGCCCCCACTCCGCTCGGCCACACCTTCGTCGA 491
   |||
67  sAspProLysValAlaGlyIleThrGlyIleThrLysGlyIleLysLeu 84
   |||
492 CGACCCCAAGGTAGCTGGCGAGCGGAAACAAAGACAAATTAAAGTTGC 541
   |||
84  rPheLysThrAlaAlaGlyLysAspValValCysIleArgSerPheGln 100
   |||
542 GGTTTAAGACTGACGACGAGAAAGATGTGGTGTCATCCGGTCTTCAG 591
   |||
101 LeuThrGlnLysAlaSerLysMetGluPheLysAlaIleGluSerValle 117
   |||
592 CTACCCAAAGCCATCAAGATGAGTTAAAGCAATTGAAACCTGCTCT 641
   |||
117 uGlnThrIleAsnProHisThrGlyGlyLysValCysLeuSerThrArg 134
   |||
642 CCAAGCTATAATATCCACACAGGGGAGAAAGTGGCTGCTCAGCTACAG 691
   |||
134 ySAlaAspMetAspArgGluIleProAlaLeuMetGlyValSerLysAla 150
   |||
692 GTGCTGACATGATAGAGAGATCTCTGCTTAATGGGTGTTCGAAGGCC 741
   |||
151 ValLeuGlnAsnValIlePheValHisGlnAspGluSerAsnTrpProle 167
   |||
742 GTACTGGAGATGTATATTTGTTGACCAGATGATCAATTCGACAT 791
   |||
167 uGlnAspProSerThrLeuLysLysPheAspAspIlePheSerIleAT 184
   |||
792 GCAGAGCCCGTCAACACTTAAGAAAGATTGAGAGACATCTTCTGCGCA 841
   |||
184 hrArgTyrThrLysAlaLeuGlnValIleLysLysLeuHisLysAspGln 200
   |||
842 CACGCTATACGAAAGCTCTGAAGTCAATAAGAACTTCACAGAGATGAA 891
   |||
201 MetGlnGluIleLysThrPheArgLeuLysLeuGlnAsnLeuGlnThrVa 217
   |||
892 ATGCAAGAGATCAAGACTTTTAGTTAAAGCTGGAGAACTTCAGACTGT 941
   |||
217 LlysAspGlnAlaHisLysLeuArgGluAsnIleAlaGlnAspGlnGlu 234
   |||
942 AAAAGACCAAGCACATTAAGTGGCGTGAATAATTCCTCAAGATCGAATA 991
   |||
234 ySerAspAlaSerLysSerGlnMetGlnGlnLysGlyLysIleCys 250
   |||
992 AGTCGATGCTCTCAAAATCTCAGATGGAGCAACTGAAGAAAGATGCTGT 1041
   |||
251 GlyThrGluArgGluIleLeuGlnMetGluThrSerLeuAspGluLeuArg 267
   |||
1042 GGTACCGAGAGAGAAATCTCTCAATGGAACAAAGTTGGATGAAGTCAAG 1091
   |||
267 ArgLeuGlnGlnGlnIleAspIleLysAlaThrGlnArgSerThrLeu 284
   |||
1092 AAGACTTGAAGGACAAATGACATCAAGGCAACAGAGAAAGTACATTAAC 1141
   |||
284 eutThrGlnGlnHisGlnLysLeuAlaAlaLeuSerGluGluAsnGluAsp 300
   |||
1142 TTACGACAGAGATGAATAAGCTGTGTCACCTTCTGAGGAAATGAAGAT 1191
   |||
301 ThrAspGlnGluLeuMetGluTrpGlnThrLysPheGlnGluArgIleAl 317
   |||
1192 ACCGATGAGGAATTAATGAAATGGCAAAATAATTTGAAGAAAGATTCGC 1241
   |||
317 aLeuLeuGlnThrLysIleSerLysLeuValArgAspMetAspAspGlu 334
   |||
1242 GTTACTGAAACAAAATCAGTAACCTTGTAAAGATGATGATGATGAAG 1291
   |||
334 lAsertyIserSerValLeuSerLysGlnAsnSerGluLeuThrHisGlu 350
   |||
1292 CATCTTATACCTCCGTTCTGTCCAAACAAATAATTCGATTAACCATGAA 1341
   |||
351 IleGlyLysLeuGlnAlaGluAlaAspAlaHisLeuThrMetLysHisG 367
   |||
1342 ATTGGAAGCTCCAGCGACGAGAGCTGATGCTCACCTGACTTGAAGATGA 1391
   |||
367 uArgAspSerAspIleLysAsnIleCysThrLysHisAsnLeuGlyPro 384
   |||
1392 ACGAGACTCAGACATATAAAATATATGACTAAACATTAATCTGGGCGG 1441
   |||
384 aLProGlnHisProPheThrAsnAspValAlaMetAsnLeuThrAsnArg 400
   |||
1442 TTCTGAAACATCCCTTTACGAATGATGTGCTATGAACCTTACAAACAG 1491
   |||
401 IleLysAlaArgLeuSerSerLeuGluAsnAspLeuAspLysLysLy 417
   |||
1492 ATTTAAACGAGACTATATAGCTTTGAGAAATGATTTGCTGGATGAAGAA 1541
   |||
1542 ATCCAAATGAAGATCAGTTAGATGTTTGTGGAAACACTATCTTAATAAA 1591
   |||
434 sAlaIleArgTyrSerGluValAspGlyGlnIleGlnSerLysIleGluSer 450
   |||
1592 ATGCTCCCTCTCTCCGAAGTGTAGTGTCAGATACAAATTAAGATGAATCC 1641
   |||
451 MetSerGlyIleLeuArgArgArgLysAspLysGluLysGluArgAspAl 467
   |||
1642 ATGTCAAGGCATTTTAACGCGAGAGAAAGATTAAGAGAAAGAACCCGATCC 1691
   |||
467 aAlaGluValGluLeuSerLysPheAsnLeuSerArgIleAspGluArgG 484
   |||
1692 TGCAGAAGTGGAGCTTCAAAATTAATCTATCCGATGATGAGAGGG 1741
   |||
484 LuArgHisMetGlnIleGluValGluArgLysThrLeuAlaLeuGlyGlu 500
   |||
1742 AGACACATATGCAAAATGAATGACAGAGAGAGACACTGCTGCGGAGAA 1791
   |||
501 ArgAspTyrAspSerIleIleSerGlnLysArgThrGluValTyrSerLe 517
   |||
1792 AGACACTATGATTTCAATTTATAGTCAGAAAGCAACAGATATAGTTT 1841
   |||
517 uGlnGlnLysIleLysValLeuLeuArgGluLysAspIleIleAsnArg 534
   |||
1842 GGAACAGAAATTAATAAGTCTTGGCGGAGAAAGATATTAATAATGAA 1891
   |||
534 sAlaAspGluArgValLysLeuGlyLeuLysLysAspAlaLeuGluSer 550
   |||
1892 ATGCTGATGAAGAGATGAAGTGGTTGGAAGGAGAGATTTGGAAGG 1941
   |||
551 SerLysAspLysLeuAsnGlnIleValAsnGlnHisLysAspLysIleLy 567
   |||
1942 AGCAAGGACAAAGCTCAATGACATAGTTATATGACATTAAGGTTAAATCAA 1991
   |||
567 sLysValLeuArgGlyArgAsnProPheGlnLysAspMetLysLysGlu 584
   |||
1992 AAAGTACTTGGGGGAGGATCCTTTGAGAGAGATGAAGAAAGAGA 2041
   |||
584 LeuAsnGlnAlaPheTrpProValAspLysGluTyrAsnGluLeuArgSer 600
   |||
2042 TCAATCAAGCCCTTTGGCTGTGGCAAGGAAATCAAAAGATTTAAGATCA 2091
   |||
601 LysSerGlnGluAlaGlnGlnLysLeuLysPheThrGlnSerLysValTh 617
   |||
2092 AAATCCCAAGGAGCAGACAAAGAGCTTAATTTACTCAGACCAAGATTAAC 2141
   |||
617 rAspAlaArgGluGlnLeuThrLysLeuArgArgAspMetAspAlaLys 634
   |||

```

2142 TGATGCTAGACAACTTGACAAAACCTCGAAGATATGATGCAAAA 2191
634 rgatgphleuaspserlyseuInserIleleuGlnIleserAlasn 650
2192 GAAGATTCCTGACTCGAACTTCAATCTATTTCACATATCTCTAAT 2241
651 VALspMetPheProLysValleuGlnsPrAlaMetAsnLysArgAspG 667
2242 GTTGACATGTTCCCAAGTTCTACAAAGCCCATGAACAAAGAGATGA 2291
667 uGlnLysArgLeuGlnuAsnPhelaAsnGlyMetArgGlnMetLeuAlap 684
2292 ACAGAAAAGATTAGAAATTCGCAATGGAATGGGGAAATGCTTGCC 2341
684 rOPheGlnHisLeuAlaArgLysAsnHisValCysProCysCysGlnArg 700
2342 CTTTGAACATTTGGCTCGAAGATCATGTATGCCCATGCTGTGAAC 2391
701 AlaPheThrProAspGlnGluAspGluPheValLysLysGlnArgMetG 717
2392 GCTTCACACCTGATGAGAGATGATGTCGTGAAGAAAACAAAGATGCA 2441
717 nasnserserThrAlaGluArgSerLysAlaLeuAlaMetGlnSerSer 734
2442 AAACCTCAAGTACTGAGAGAGATCTAAAGCTCGCAATGGAATCTCA 2491
734 snAlaGlnAlaLeuPheGlnGlnLeuAspLysLeuArgThrIleArgAsp 750
2492 ATGCTGAACCTTTTTCAGCATGATGAACTTCGCACTATCTATGAT 2541
751 AlaTyrValLysLeuValGlnGlnThrIleProLeuAlaGlnLysAsnLe 767
2542 GCTATGTCAGCTGCTAGAGAAACCATCTCTAGCAGAGAAAAACT 2591
767 uAsnGlnHisLeuAlaAspGlnSerGlnLysAlaGlnAlaPheAspSer 784
2592 GAATCAACATTTGGCGGATGAAGTCAAGAGCGCGCTTTGATGATC 2641
784 euLeuGlnValLeuAlaHisValGlnMetAspArgAspAlaValGlnAla 800
2642 TTTTGGGCTTCTTGCCCATGTTCAATGACAGCGATGCGATGGAAGCC 2691
801 LeuLeuGlnProThrAspThrIleAspArgHisValHisGlnIleGln 817
2692 TTATTCACACCACATGATCTATGACAGCGATGATCACTAAATTCACA 2741
817 nLeuValLysGlnValGlnAspLeuGlnLysArgAlaLeuAspSerSerGly 834
2742 GCTAGTCAAGAGATGAGATCTGAATATGCACTTGATTTAGTGCC 2791
834 rGlnValLysSerLeuGlnGlnIleGlnLeuGlnLeuAsnPhelaGln 850
2792 GAGGTTCAGTCTTTCGAGAAATTCACACTGACGCTGAACTTCTGAC 2841
851 ArgThrArgAspThrLeuIleValGlnValAspAspLeuArgAspGln 867
2842 AGAACACAGGACACATGATTCGCAAGTGCATGATCTTAGACATCAACA 2891
867 sarGMetLeuAsnGlnAspMetSerSerAlaGlnValArgTrpHisAsn 884
2892 TAGAAATGCTAAATGAAGATATGCAAGTGTCTCAGGTGATGAGACATG 2941
884 laArgGlnGlnLysValLysAlaSerSerIleLeuGlnArgPheGlnLys 900
2942 CTGGGAGAGAAAGTCAAGCTTTCGATATTCGAAAGATTCACAAA 2991
901 SerGlnGlnLysLeuValLeuLeuAlaGlnGlnLysGlnGlnIleVal 917
2992 TCTGAAGAGAAATGCTTTCAGCTGAGAGAAAACAAACATGATGT 3041
917 lGlnLysLysLeuLeuGlnGlnLysLeuAspProLeuSerLysGlnLys 934
3042 AGAAAGAACTTTTAGAAGATCTCTTGATCCATGTCACAAAGAGAAAG 3091
934 lSerLeuLeuGlnGlnLysArgAsnAlaLeuLysGlnLysLeuAspGln 950
3092 AGACCTTGCTGACAGATTAATGCTTTGACAAAGCTGATGAAGAG 3141
951 TyrHisGlnLeuAlaGlnArgLysArgGlnPheGlnGlnGlnLeuAsp 967
3142 TATCATACACTTCAGAAAGAAAGAGAGGAGTCCAGCAAGATCTGATGC 3191
967 AluGlnArgLeuAsnMetLysIleLysGlnLysLeuAspSerLysLys 984
3192 TCTTGAGAACTTAATATGAAGATTAAGAGGATCTGTTCCAGAAAA 3241
984 snGlnLysLeuLysGlnGlnGlnArgHisValLeuCysHisSerGln 1000
3242 ACAGAAAGCTTAAGAAATTCAGAGAGACATCTTTCCTGATCTCAG 3291
1001 LeuGlnSerCysMetAlaLysGlnGlnArgIleSerAlaGlnLeuAsn 1017
3292 TTTACAGATTCATGACCAAAACAGAAAGATATAGCTGATTAACAA 3341
1017 sSerLysGlnLeuLeuGlnGlnGlnGlnGlnGlnLysArgAsnIleAsp 1034
3342 GAGCAAGAAGTACTGACAGGCGCAGGCGGCTTGAAGAAACATGATG 3391
1034 sPAsnLeuLysTyrArgLysThrLysAlaAspValGlnGlnLeuThrArg 1050
3392 ACAATCTCAAGTACAGAGAAACAAAGCTGATGTCGAACATCTACT 3441
1051 AspIleGlnSerLeuGlnGlnArgLeuLeuSerIleGlnSerLeuSerAl 1067
3442 GATATGATCACTTGAAGAAAGCTGCTTCAATAGATAGCTGTGCTGC 3491
1067 AlIleGlnAlaAspLeuLysArgHisSerGlnGlnLysGlnValGlnAsn 1084
3492 TATAGAAAGCTATGTAAGCGCATCTCTCAAGAAAAAGAGGCTTAAT 3541
1084 erGlnPheAsnArgTrpGlnGlnLysLeuSerValTyrGlnSerAsnIle 1100
3542 CAGAAATTTAAGAGTGGCAAGAGAACCTTCTTTTATCAAAATATT 3591
1101 SerLysHisLysGlnGlnLeuLysLeuSerGlnTyrLysAspIleGln 1117
3592 TCAAGACCAACAAAGAGCTTAAGCTTCACAGTACAGATATCAGAA 3641
1117 sarGlyThrAsnGlnPheLeuGlnLeuLysThrThrGlnMetAlaAsn 1134
3642 GCGATATACTAAATCAATTTCTCCAGCTTAAGACAACTGAATGGCAACA 3691
1134 yAspLeuAspArgTyrTyrThrAlaLeuAspLysAlaLeuMetArgPhe 1150
3692 AGGACTTGACAGATTAATATCTGCTTAAACAGGCTCTTAATGCGGTT 3741
1151 HisSerMetLysMetGlnGlnIleAsnLysIleLysGlnLeuTrpGln 1167
3742 CACAGCATGAAGATGAGAGATTAATTAATTAATCAAGAACTGTGCA 3791
1167 nGlnThrTyrArgGlnAspIleAspTyrIleSerIleAsnSerAsp 1184
3792 ACAGACATACAGAGCCAGATATGATTAATTAATTAATTAATTAAT 3841
1184 erGlnGlnAlaGlnLysThrArgSerTyrSerTyrArgValValMetGlnThr 1200
3842 CTGAGGCTGCGGACATGATACAGCTACAGCCGCTGTTGATACAACT 3891
1201 GlyAspAlaGlnLeuGlnMetArgGlyArgCysSerAlaGlnLysVal 1217
3892 GGTATGCTGAGCTGGAATTCGAGGCGCTGAGTCTGTGTAAGAGGT 3941
1217 lLeuAlaSerLeuIleIleArgLeuAlaLeuAlaGlnLysPheCysLeu 1234
3942 TCTTGCTCTCTTATTAATCAAGATGACATTCGCGAAACTTTCGCTGA 3991

1234 sncysgllleleualaleuaspplubprothrrhthasleuaspplpro 1250
 |||||
 3992 ACTGCGGATATTTGGCTTTGGATGAGCAACTAGCAATGATGAGGCA 4041
 |||||
 1251 AsnIaGusSerleuAlaAlaAlaLeuLeuArglleMetGluAlaAagly 1267
 |||||
 4042 AATGCAAGAGTCTTGGCTGCGCCCTGTGGATATATGAGACAGGAA 4091
 |||||
 1267 sglYnglInuAnpHeGlnleuIleValIleThrHisaspGluArgphea 1284
 |||||
 4092 AGGCGAGAGAACTCCAGTGTATCATCATCATGATGAGAGATTGG 4141
 |||||
 1284 lahlaleuilegllnglnargglnleuAlaGlytyrtyrArgValSer 1300
 |||||
 4142 CCCATCTTATGCTGTCAGAGGAGCTGCTGAGAGTACTATGAGTCC 4191
 |||||
 1301 LysaSpGluAnGlnHisSerlleleGluSerGlnIuilephea 1316
 |||||
 4192 AAGGATGAGAAACGACACGATATTTGATCCAGAGATATTGGAC 4239
 |||||

seq_name: gb_p12:AF168748

seq_documentation_block:
 LOCUS AF168748 4305 bp mRNA 01-MAR-2000
 DEFINITION Arabidopsis thaliana DNA repair-recombination protein (RAD50) mRNA,
 complete cds.
 ACCESSION AF168748
 VERSION AF168748.1 GI:7110147
 KEYWORDS
 ORGANISM
 Arabidopsis thaliana
 thale cress.
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 4305)
 Gallego, M.E., Nagpal, P., Quatrano, R. and White, C.I.
 The RAD50 homolog of Arabidopsis
 Unpublished
 JOURNAL
 2 (bases 1 to 4305)
 Gallego, M.E., Nagpal, P., Quatrano, R. and White, C.I.
 Direct Submission
 TITLE
 Submitted (13-JUL-1999) UMR 6547 - CNRS, Université Blaise Pascal,
 24, Ave. des Landais, Aubiere 63170, France
 JOURNAL
 24, Ave. des Landais, Aubiere 63170, France

FEATURES
 source
 1. 4305
 /organism="Arabidopsis thaliana"
 /cultivar="Columbia"
 /db_xref="taxon:3702"
 /chromosome="II"
 /map="near TENS"
 /note="Cloned from cell suspension culture"
 1. 4305
 /gene="RAD50"
 146. 4096
 /note="similar to yeast RAD50"
 /codon_start=1
 /product="DNA repair-recombination protein"
 /protein_id="AF36810.1"
 /db_xref="GI:7110148"
 /translation="NSTVDMMLKGRSPDENKNVTPRPRTILVANGAKRTTI
 ECTKSTGELPPNANSGHSFIDPKVAGETETKAOIKLFEKTAAGKDVVCTISFOIT
 OKAKMEYKAIESVLOTINPTGEKVCLSADMDREIPALNGVSALILENIVHQ
 DESNMPLODPTLKKKFDIFSATRYKALEVTKLHKDOAEIKTFKLEMLNLOTIK
 DAAKLAESIAOQERESSKVOMLLETSVQKVAEHNKEMMLKDLRKLOQVSIK
 TARSSTLFEKQOQVAYALPEENEDTIELEKMSKEPERALALGTKRKREMAVORE
 TTISLHATNTMLESKIQTEAEMMLKLNROSTIOIFHYHYNLGNPSPFSE
 VVNLNTRIKSRGELEMDLDRKRSNETALSTAMPDQMDANRMSISLEKAKDEI
 KMGSKRIEKEKEIERSFEFISTVDKQDDEKQOVLEKRTKQNSGSKESKE
 OKOHEIYSLSEKIKTLNREDDVAGDAEDVKLSLKTDEENLKKKKIIDECKRI
 RGVLKRLPEKMDKREIVOLRSIEREYDLSLSKSEAKEVNMLQMKLOEVNNSIF
 KHNKDESRKRYESTKLOAKOSVITDAYPKLLESAKKDRDRKREYNANAMRQWF
 EPPEKRAKROHSCPCCEBRSPTADEASFIKKQVKASSTGEHKLALAVESSNADSVFO

gene
 CDS

OLDKRAVFEESKLTTEITIPLEAKETLOEHTEELCOKSEALDVLGISAOIKADKDSI
 BALVPLENADRIPOEIVSYOKOIEDLEKLPDRGLKYTHEISSELSQSKDL
 HGELEKLPDOYMEYERDISCOARWHAAREKAKANILROVTKAEDELRLAEKSO
 LDLDVRYTLBALGPILSKEREDELSDYNDMKIRNOBYELAKKNYQOEVALKAS
 YKINBYHDKKEERLDDIOEKORLSDSQOSEAKNELAGELNRKDLNRQDOLR
 NIEDLNRTTAKVEELRELESLEOILNNGIAVAEALVKILFRERELSELNR
 CRGTVSESSISKNRVELKOAYDIDKRFHQDLOIKTEEMARKDIDRYNALDKA
 LMRFHMKKEEINKIIRLEMOQTYGODMDYRIHSDSEAGTSRSYKVLMOGTDE
 LEMRGCASGOKVLAISLIRLALAEFPCLNGIILADEPTTLDGPNSESLAGALLRI
 MEDRGOERFOLIVTTHDERFRQMTGQROHAKERYIRVAKDMQHSIIENQELFD"

BASE COUNT 1544 a 733 c 999 g 1029 t
 ORIGIN

alignment_scores:
 Quality: 4288.00 Length: 1316
 Ratio: 3.771 Gaps: 0
 Percent Similarity: 86.398 Percent Identity: 63.450

alignment_block:
 US-09-538-396-2 x AF168748

Align seq 1/1 to: AF168748 from: 1 to: 4305

1 MetSerThrValAspLysMetLeuIleLysGlyIleArgSerPheAsp 17
 |||||
 146 ATGAGTACGGTCGATAAATTTGATCATCAGGCTATAGAACCTCGACCC 195
 |||||
 17 oaspAsnLysAsnValIleThrPhePheLysProLeuThrLeuIleValG 34
 |||||
 196 AGAAATATAGAACGTTGTTACTTCTTCAGACCTTAAACCTAATCGTCG 245
 |||||
 34 LyrProAsnGlyAlaGlyLysThrThrIleLeuGlyLysLeuLysSer 50
 |||||
 246 GCGCTAACGAGCCGGAACAACTAGATATAGATGATTAAGGTGTCT 295
 |||||
 51 CysThrGlyGluLeuProProAsnSerArgSerGlyHisThrPheValH 67
 |||||
 296 TGTACCGAGAGTTCGCCACGACGCGAGTCTGCTACCTTACCTTCA 345
 |||||
 346 TGACCTTAAGTGTGTGTGAGAAACGAGACAAAGACAAATAAAGCTGA 395
 |||||
 67 saspProLysValAlaGlyLysThrThrIleLysGlnIleLysLeuA 84
 |||||
 84 tgpheLysThrAlaAlaGlyLysAspValAlaCysIleArgSerPheH 100
 |||||
 396 GATTCACACAGCTGCGACGAGAGTGTGTATATATAGGTCATTTCAG 445
 |||||
 101 LeuThrGlnLysAlaSerLysMetGluPheLysAlaIleGluSerVal 117
 |||||
 446 TTGACGCAAAAGCTTCAAGATGAGTACCAAGCTATTGAGAGTGTCT 495
 |||||
 117 uGlnThrIleAsnProHisThrGlyLysValCysLeuSerTyrArgC 134
 |||||
 496 TCAACCATTAATCCTCACACTGAGAGAAAGTANGTCTAGCTACAGAT 545
 |||||
 134 ysaIaaspMetaspArgGluIleProAlaLeuMetGlyValSerLysAla 150
 |||||
 546 GCGCTGACATGATAGAGATGAGATACACAGCTTGTATGGGTGTCAAAAGCT 595
 |||||
 151 ValLeuGluAsnValIlePheValHisGlnAspGluSerAsnTrpPro 167
 |||||
 596 ATTTAGAAAGAGCTATATTTGTGCATACAGTATGAGTAACTGGCCACT 645
 |||||
 167 uGlnAspProSerThrLeuLysLysLysPheAspIlePheSerAlaT 184
 |||||
 646 GCAAGATCTCTTACTTGAAGAAAGAGTATGATATTTTCTCAGCA 695
 |||||
 184 hrArgTyrThrLysAlaLeuGluValIleLysLysLeuHisLysAspGln 200
 |||||
 696 CCAGATATACAAAGGCTCTGGAAGTTATTAGAAAGCTCACAAAGATCAA 745
 |||||
 201 MetGlnGluIleLysThrPheArgLeuLysLeuGluAsnLeuGlnThrVa 217

```
|||||.....
746 GCTAGGAGATAAAACTTCAAGTTAAATGGAAAAATCGACAGCTT 795
217 LlysAspGlnAlaHisLysLeuArgGluAsnIleAlaGlnAspGlnGlu 234
796 GAAAGACGACCTTATAGCTTCGTGAGACATTCCTCAAGATCAAGAAA 845
234 ySerAspAlaSerLysSerGlnMetGlnLeuLysGlnLysIleCys 250
846 GGACAGAAATCTTCAAAAGTTGAGATGTGAACTGGAACCGCTGTACG 895
251 GlyThrGluArgGluIleLeuGlnMetGluThrSerLeuAspGlnLeuAr 267
896 AAAGTTGAGCTGAGATTCATTAATAGAAATGATGTTAAAGATCTGAG 945
267 GArgLeuGlnGlnIleAspIleLysAlaThrGluArgSerThrLeu 284
946 AAAGCTCAGAGCAACAGTATCAATAAACTCGGAAAGAGACACTTGT 995
284 eUTHrGlnGlnHisGluLysLeuAlaAlaLeuSerGluGluAsnGluAsp 300
996 TTAAGGACACAGACAGACAGTATCGACATTCCTGAGGAAATAGAAAT 1045
301 ThrAspGluGluLeuMetGluThrGlnThrLysPheGluGluArgIleAl 317
1046 ACTATTGGAATTTGAAAGATGAAAGCAAGTTTGAAGAAAGACTTGC 1095
317 AleuLeuGluThrLysIleSerLysLeuValArgAspMetAspAspGlu 334
1096 TCTTTTGGGAGCAAAATTCGAAAAATGAGAGAGATGGTTGACACAG 1145
334 LaSerTySerSerValLeuSerLysGlnAsnSerGluLeuThrHisGlu 350
1146 AAACATCAATTTCTCGCTTCAAAATGCAAAAACTAACTATATGCTGCG 1195
351 IleGlyLysLeuGlnAlaGlnIleAspAlaHisLeuThrMetLysHisG 367
1196 ATCAGCAACCTACAAACAGAGACTGAAAGCTCATATGCTGTAAGAAATGA 1245
367 uArgAspSerAspIleLysAsnIleCysThrLysHisAsnLeuGlyPro 384
1246 AAGAGATTTTACCATTCCAGACATTTTTCATTCATTAATTTGAAATG 1295
384 aProGlnHisProPheThrAsnAspValAlaMetLysLeuThrAsnArg 400
1296 TTCCTGATACCCCTTCAGTACAGTAGTTTGAATCTCACAAATTCGA 1345
401 IleLysAlaArgLeuSerSerLeuGlnAsnAspLeuLeuAspLysLys 417
1346 ATAAAAATCAAGACTGGGTAACCTGAGATGATTGCTGATTAAGAGAA 1395
417 sSerAsnGluAspGlnLeuAspValLeuThrPlysHisTyLeuLysIle 434
1396 ATCAAAATGAAAGCTCTTAAGTACGGCTTGCGATGCTATATGAGAGCTA 1445
434 snAlaArgTySerGluValAspGlyGlnIleGlnSerLysIleGluSer 450
1446 ATGACCGTTGGAAGAACATCGAAGCTCAGAAACGACGCTAAAGATGAAATC 1495
451 MetSerGlyIleLeuArgArgArgLysAspLysGluLysGluArgAspAl 467
1496 AAGATGGGATATCAAAACGATAGAAAGAAAGAAATTAAGCGCTGATTC 1545
467 aAlaGluValGluLeuSerLysPheAsnLeuSerArgIleAspGluArg 484
1546 ATTTGAATTCGAAATTCCTACCTGATGTTAAACAAACGATGAAAGAG 1595
484 LuArgHisMetGlnIleGluValGluArgLysThrLeuAlaLeuGlyGlu 500
1596 AGAAACAGAGTGAAGTTCAGCTTGAGAAAGACAAACAGAAATTCGAA 1645
501 ArgAspArgSerSerIleIleSerGlnLysArgThrGluValTySerLe 517
|||||.....

1646 AGGGATTGAATCCAAATAGAACAGAAACGATGAAATATACAGCTT 1695
517 uGluGlnLysIleLysValLeuLeuArgGluLysAspIleIleAsnArg 534
1696 GGAGCATTAATCAAGACACATCGAATCGAAGAAAGAGACTTTAGCTGTG 1745
534 snAlaAspGluArgValLysLeuGlyLeuLysLysAspAlaLeuGluSer 550
1746 ATGCTGAGATAGAGTGAATATGTTCTTAAAGAAACGCAACAGAGATAT 1795
551 SerLysAspLysLeuAsnGluIleValAsnGlnHisLysAspLysIle 567
1796 CTTAAGAAAGAACACAAAAGTAAATAGATGATGCAAGGATCGAATTAG 1845
567 sLysValLeuArgGlyArgAsnProPheGlnLysAspMetLysGluI 584
1846 AGGAGTGGCGAAGGGAGATTTCCACCCGAGAGATATGAAAGAGGAAA 1895
584 LeAsnGlnAlaPheThrProValAspLysGluTyArgAsnGluLeuArgSer 600
1896 TAGTTCAGCTTTGAGGTCATTTGAAGGAAATAGATGATGCTCAGTTTG 1945
601 LysSerGlnGluAlaGluGlnIleuLysPheThrGlnSerLysValTh 617
1946 AATCCCGAAGACAGAAAGAGGTCAACATGTCAGATGAAAGATACA 1995
617 rAspAlaArgGluGlnLeuThrLysLeuArgArgAspMetAspAlaLys 634
1996 AGAATGAACCAATAGGCTTTCAAGCATTAACAGATTAACAGATACAGA 2045
634 rArgPheLeuAspSerLysLeuGlnSerIleLeuGlnIleSerAlaAsn 650
2046 AGAGATACATTTGATGCTCAAGCTTCAGGCTTAAACAGAAATCAGTTAC 2095
651 ValAspMetPheProLysValLeuGlnAspAlaMetAsnLysArgAsp 667
2096 ATGATGCTTATCCCAATTTGTGGAGTGGCTTAAGCAAGAAAGAAATGA 2145
667 uGlnLysArgLeuGlnAsnPheLysAsnGluMetArgGluMetLeuAla 684
2146 CCGTAAACGCGAATATATACATGGCAATGGATATGAGCTCAAGATTTGAA 2195
684 rPheGlnHisLeuAlaArgLysAsnHisValCysArgProCysGluArg 700
2196 CTTTGAAGAAACGACACATCGAAGACATTTCTGTGCTGCTGCGAGGCC 2245
701 AlaPheThrProArgLeuGluAspGluPheValLysLysGlnArgMet 717
2246 TCTTTAAGCTGATGAGAGGCTACCTTATTAAGAACCAAGGGGTAA 2295
717 naAsnSerSerThrAlaGlnArgSerLysAlaLeuAlaMetGluSerSer 734
2296 GGCTTCAAGCAGAGGAGGACATTTAAAGCATTAAGCGAGGATCTTCCA 2345
734 snAlaGluAlaLeuPheGlnGlnLeuAspLysLeuArgThrIleTyAsp 750
2345 ATGCTGACTGTGATTTCAACACATGACAAAGATAGCGGAGTTTGA 2395
751 AlaTyValLysLeuValGluGluThrIleProLeuAlaGluLysAsnIle 767
2396 GAATACAGAACTACTACTGAAATATATCTCTGCTGAGAAACCTTT 2445
767 uAsnGlnHisLeuAlaAspLysSerGlnLysValGlnAlaPheAspArg 784
2446 GCAAGACCAATACAGAAAGCTGGTCAAGTGTGAAGCTCTTGATGATG 2495
784 eUleuGlyValLeuAlaHisValGlnMetAspArgArgAspAlaValGluAla 800
2496 TATTAGCAATCTCGCAGACATAAAGCTGCAAGAGCTGATTAAGCA 2545
801 LeuLeuGlnProThrAspThrIleAspArgHisValHisGluIleGlnI 817
2546 CTGCTGACGCCCTCGAAATATGCTGACAGATCTTTCAGGAAATCGTATC 2595
```

817 nleuValLysgluValgluAspLeuIngluTyrAlaLeuAspSerSerGlyA 834
 2596 CTACCAAAAGCAAAATGAGATCTTGAAATTAACCTTGATTTTAGGGGTC 2645
 834 rglValLysSerLeuIngluInleuInleuInleuAsnPhleuIn 850
 2646 TTGGGGTTAAACATGAAAGAAATCCATGAGAAATTAAGTAGCCTCCAA 2695
 851 ArgThrArgAspThrLeuLeuValGluValAspAspLeuArgAspGlnH 867
 2696 AGCAGCAAGATTAATATGACAGCGTAGCGAGCAACTTGAAGATGACCA 2745
 867 sarMetLeuAsnGluAspMetSerSerAlaGlnValArgThrHisAsn 884
 2746 GATTACTTACGACGATATATATCATGCTCCAGCCAGCCAGATGATGACG 2795
 884 laArgGluGluValLysAlaSerSerLeuGluArgPheGlnLys 900
 2796 TAAGGAGAGAAAGCCAGCCAGCCAAATTTGACGGAGCTTACAAAG 2845
 901 SerGluGluGluValLeuValLeuAlaGluGluLysGluInleuLeu 917
 2846 GCAGAAAGAGACCTGAGCGCTTACCGAGGAGAAAGTCAACTTGACCT 2895
 917 GluLysLysLeuLeuGluInleuSerLeuAspProLeuSerLysGluLysG 934
 2896 CGATGTGAAGATTTGACCGAGCGCTGGCCCTTTATCCAAAGGAAAG 2945
 934 luserLeuLeuGlnGluTyrAsnAlaLeuLysGlnLysLeuAspGluGlu 950
 2946 AGCATTTCTAGTACACATATGATATGAAGATTAAGCATTAACCAAG 2995
 951 TyrHisGluLeuAlaGluArgLysArgGluPheGlnGluLeuAspAl 967
 2996 TATGAAGAACTGGCTGAGAAAGAAAGAAACCAACAGAGAAAGTTGAGGC 3045
 967 aleuGluArgLeuAsnMetLysLysLysGlyTyrLeuAspSerLysLys 984
 3046 ATTACTTAAGGCCAGTTTAAATTAATGAATATCATGATCTGAAGAAAG 3095
 984 snGluLysLeuLysGluLeuGlnArgHisValLeuCysHisSerGln 1000
 3096 GGGAGCGCTTGAATGATATTCAGAAAGAAAGCAGCGTTATCTGATTCAG 3145
 1001 leuGlnSerCysMetAlaLysGlnGlnArgLysSerAlaGluLeuAsn 1017
 3146 CTTCAGACTGTGAAGCTAAGAAAGATGAATGACTGAGGTAAGTGAACG 3195
 1017 sSerLysGluLeuLeuGlnGlnArgLysGlnLeuLysArgAsnLeuAsp 1034
 3196 AAATTAAGAACTGATGAGGAAATCAAGATCACTGAAGAAATATGAGG 3245
 1034 spAsnLeuLysTyrArgLysThrLysAlaAspValGluGlnLeuThrArg 1050
 3246 ATTAATTTAACTACCGGCAACAAAGAAAGAAAGTGAAGGAGACTTACAG 3295
 1051 AspLeuSerLeuGluGluArgLeuLeuSerLeuGlySerLeuSerAl 1067
 3296 GAGATTGATCATTTGAGAGCAAAATATTTGATTTGAGGATGCGCCG 3345
 1067 alLeuLysAlaAspLeuLysArgHisSerGlnGlnLysGluArgLeuAsn 1084
 3346 AGTTTAACCTGAATTTGAAGATATTGCGGGAAAGGAGAAAGACTTCTTT 3395
 1084 ergLysPheAsnArgTrpGlnGluThrLeuSerValTyrGlnSerAsnLe 1100
 3396 CAGACTGTAACCGGTGCTGAGAACAGTCTCTTTACGAGACGAGTAT 3445
 1101 SerLysHisLysGlnGluLysLeuSerGlnTyrLysAspLeuGlyL 1117
 3446 TCAAGAACAGAGTGGAGCTAAACAGCAATATCAAGAGACATGATTA 3495

1117 sarGlyThrAsnGlnPheLeuGlnLeuLysThrThrGluMetAlaAsn 1134
 3496 GCGGCACTTTGATCACTGATCCAGCTAAAGAACATGAATGCAATA 3545
 1134 ysAspLeuAspArgTyrTyrThrAlaLeuAspLysAlaLeuMetArgPhe 1150
 3546 AAGACTGTGATGATCACTACATGCCCTTACAAAGCAATATGCGCTTC 3595
 1151 HisSerMetLysMetGluGluLeuAsnLysLleLysGlnLeuTyrPgl 1167
 3596 CACCAATGAAATGAGAAATTAATTAAGATTATTAAGGAGACTGTGCA 3645
 1167 ngInThrTyrArgGlyGlnAspIleAspTyrLysSerLysAsnSerAsp 1184
 3646 GCAGACATACAGAGCTCAGATATGATTCATTCAGATACACACAT 3695
 1184 ergLysAlaGlyThrArgSerTyrSerTyrArgValValMetGlnThr 1200
 3696 CCGAAGGTGAGCGCCCGCTTACAGCTACAGAGTTCATTATGACACT 3745
 1201 GlyAspAlaGluLeuGluMetArgGlyArgCysSerLeuGlyGlnLys 1217
 3746 GGTGATACAGAACTTGAATGAGAGAGATGACAGTCAGGTCAAAGGT 3795
 1217 lLeuAlaSerLeuLleLeuArgLeuAlaLeuAlaGluThrPheCysLeu 1234
 3796 TCTTGCTTCATGATATTAATGATGCTTACCGGACATTTGCTTA 3845
 1234 snCysGlyLleLeuAlaLeuAspGluProThrThrAsnLeuAspGlyPro 1250
 3846 ACTGTGAATATTTAGCCCTGATGAGCCAACTACAAACCTGAGAGGCC 3895
 1251 AsnAlaGluSerLeuAlaAlaAlaLeuLeuArgLysMetGluAlaArgLys 1267
 3896 AATTCAAATCTCTTGAAGAGCTCTTCTTGAATCATGAGAGACAGAA 3945
 1267 sGlyGlnGluAsnPheGlnLeuLysLysLysLysLysLysLysLys 1284
 3946 GGGCCAGAGAAATTTCCACATGATGATTCATCCAGATGAAAGCTTG 3995
 1284 lAlaHisLeuLleGlyGlnArgGlnLeuAlaGluLysTyrTyrArgValSer 1300
 3996 CTCAAATGATGGCCAAAGCAACATGCTGAAATATTAACCGGCTGCA 4045
 1301 LysAspGluAsnGlnHisSerLleLysGlnSerGlnLysLysPheAsp 1316
 4046 AAAGATGATATGCAACAGCATATGAGGCCCAAGAGATCTTTGAT 4093
 seq_name: gb_p11:AC006223
 seq_documentation_block:
 LOCUS AC006223 100567 bp DNA PLN 05-APR-2000
 DEFINITION Arabidopsis thaliana chromosome II section 180 of 255 of the
 complete sequence. Sequence from clones F20M17, F22D22.
 ACCESSION AC006223 AE002093
 VERSION AC006223.3 GI:6598534
 KEYWORDS HTG.
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE 1 (bases 1 to 100567)
 Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
 Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E.,
 Feldguy T.V., Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M.,
 Koo H., Moffat K.S., Cronin L.B., Shen M., Vanaken S.E., Umayam L.,
 Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,
 Goodman H.M., Somerville C.R., Copenhaver G.P., Pruss D.,
 Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M. and
 Venter J.C.
 TITLE Sequence and analysis of chromosome 2 of the plant Arabidopsis
 thaliana

alignment_scores:

Quality: 3050.00 Length: 2727
 Ratio: 2.661 Gaps: 45
 Percent Similarity: 42.024 Percent Identity: 30.436

alignment_block:

us-09-538-396-2 x AC006223 ..

Align seg 1/1 to: AC006223 from: 1 to: 100567

```

1 MetSerThrValAspLysMetLeuIleLysGlyIleArgSerPheAspPr 17
|||||
17700 ATGAGTACGGTCGATTAATGTTGATCAAGGATATAAGAGCTTCGACCC 17749

17 cnspsnlysnvalliethrphelysPheLeuThrIleValG 34
|||||
17750 AGAAATAGAAAGCGTTGTTACTTCTTCAGACCTTAACTTAACGTCGCG 17799

34 LYPProAsnGlyAlaGlyLysThr..... 41
|||||
17800 GCGCTACGAGACCGGAAAGCTTAAGTCTTAAACCTTTTCTCTTC 17849

41 ..... 41

17850 TCCTCAGTTCACTCCCATTTCTTCGCTCCGTAATTGAAGTTGTGAC 17899

42 ..... 52
|||||
17900 TATTTGCGATTTTTCAGACGATTAAGAGTGTAAAGGTTGTTGTA 17949

52 hrGlyLeuProPheAsnSerArgSerGlyHisThrPheValHisAsp 68
|||||
17950 CCGGAGAGTTCGACCGAAGCGAGAGTCTGTCATAGCTTCATTCATGAC 17999

69 ProLysVal..Ala..... 72
|||||
18000 CCTAAGTAGTACTCTAGTTAGATTCCTCAATTTACCTAATTTCAATGTT 18049

72 ..... 72

```

```

18911..19025,19249..19477,19659..19837,19925..20095,
20187..20291,20381..20488,20588..20740,20875..20978,
21144..21271,21551..21665,21785..21914,22033..22232,
22429..22637,22749..22950,23535..23704,23764..24100,
24481..24606,24692..24744,24855..25005,25081..25143,
25254..25387,25590..25722)
/gene="At2g31970"
/codon_start=1
/product="putative RAD50 DNA repair protein"
/protein_id="RAD15407.1"
/db_xref="GI:4263721"
/translation="MSTVDKMLIKGIRSPENKNVYFRRLLIVGANGAKTII
ECLAKTSCTELPNARSCHSFTHDRKVAETEDKQIKRFTAKGVCCRSFQLT
OKASKMEYKAIESVLQFINPHGEKVCISYRCADNRREIPALMGSKALLENITFHO
DAENMPLQDPSTLKKKFDIFSATRYTALLEYIKLHKQAOEIKTKFLKLEQLTK
DAAYKLREIAQOERTESSKQOMLETVOVDVAENKEMKMLDKRLKLDQVSIK
TAEKSTLFEKQOYALPEENEDTIEELKEMSKFEERLALGTRIKRMERMVTE
TTTSSAHNAKRYMLEISKLOTEAEAHMLKKNRSTIONIPHYMLGVSPPESTE
VYVNLNRIKSRGELDELDDKKSNETALSTANDCYDANDRMKSTIAOKRANDEI
KMGISRIEKEIERDSFEFEISTVDVKOTDEREKOVVELEKTKQNSERGEKTE
OKOHEYSIEIEEDLSLKSREAEKVNLMOKIOWNSLFEKHNDESRKRYIESK
EIVQALRSIEREDLSLKSREAEKVNLMOKIOWNSLFEKHNDESRKRYIESK
IQAIOESVTIDAYPKLESADKRDREKREVMANGMOMPEPFRARORESCCC
ERSTPDEASPKKORVKAASSTGHEILKALAYSSNADSVFOLDKLRVFEESLT
TEIPLAEKTLQEHTELOKSEBALDVIGISQIRADDDSTALVQPLENDRITQE
IYSYQOIEDLEKLOFRLGVKTMEEIQSESSLOSXDKLGELEKLDQITMER
DISCLOARHAAVEERAKAANLIDVTKAEEDELAERKSOQLDVKYLTEGLS
KEKEQLSDYNDKIRRNQYERELAKRNRYOQVALLKASIKINDCTRYDLKLG
ERLDDIOEKORISQLOSCARKNELAGELNNKRNODLRINIDNLNRYPTK
AKVEETRTRESLEEDILNIGCALAYELIVKILREBELSLNCRGTVSYESSI
SKRVELKQKQIDKIDKHFDDLIQKTTMANKDIDRYNALDKRLMFRHTKMEI
NKIRLMQOYTRGQMDYIRHSDEGAGTNSYKVLMTQSDTELEMGRCSAQOK

```

```

18050 TCCTGTGTGATGATGNGAATGCAGATTCGATTTGCCCTATAGTAG 18099
73 .....GlyGut 75
18100 TATTCAGATTGAATTCATTGTTCTCCACTATATAGGTTGCTGAGAGAA 18149
75 hrGluThrLysGlyLniIleLysLeuArgPheLysThrAlaAlaGlyLys 91
|||||
18150 CCGAGACAAAGACAAATTAAGCTGAGATTCAGACAGCTCCAGGAGAG 18199
92 AspValValCysIleArgSerPheGlnLeuThrGlnLysAlaSerLys 108
|||||
18200 GATGTGATGTATATAGGTATTCAGTTGACGCAAAAGGCTCAAGAT 18249
108 tGluPheLysAlaIleGluSerValLeuGlnThrIleAsnProHisArg 125
|||||
18250 GGAATACAAAGCTATATGAGATGTCCTCAACCATTAATCTCAGACTG 18299
125 LysLui..... 126
|||||
18300 GAGAGGTTGCACCTTATCTCTGTTATTATCTTCATTGTTGTTCTCTC 18349
126 ..... 126
18350 TCCTTTTTCATGATGATATAGATCTGATTTTGTGTAAGGTCTC 18399
126 ..... 126
18400 GAGCTAAGTGTAGACTATCTATATTCCTACAAAGTTACAACGTTTGTT 18449
126 ..... 126
18450 TGTCTGATATAGATGATTTGATTAACCTGTTTGTGAGAAAGCTTGA 18499
126 ..... 126
18500 ACGCAGCCTCATTTGAGGTTTCTAATGTTGTTTGTATTATTCCTCC 18549
126 ..... 126
18550 ATTTTTCATGATGATTCATCATGATCATTTCTTTCTTATTTTGTCC 18599
127 .....Lys..ValCysLeuSerTyrArgCysAlaAs 136
18600 ACATACATGTCMAATTCGAGAAAGTATGTCACGTCACAGTGGCGTGA 18649
136 pMetAspArgGluIleProAlaLeuMetGlyValSerLysAlaValLeuG 153
|||||
18650 CATGATAGGAGATACACAGCTTGTATGGTGTGTCCAAAGCTATTTAG 18699
153 LuAsnValIlePheValHisGlnAspGluSerAsnTrpProLeuGlnAsp 169
|||||
18700 AAACGTCATATTTGTGCATCAGATGAGTCAACTGCGCCACTGCAAGAT 18749
170 ProSerThrLeuLysLysLysPheAspAspIlePheSerAlaThr..... 184
18750 CCTTCTACTTTGAAAAAGAGTTGATGATATTTTCTGAGCCACGAGTA 18799
184 ..... 184
18800 GTTTCTGGCTACCCACCAACATTTCTTAATCATCATGTTCTGCTGT 18849
184 ..... 184
18850 ATTTGTTGTAGAGAAACCGAATCCAGAAATGTAAACATGTGTGTTA 18899
185 .....ArgTyrThrLysAlaLeuGlnValIleLysLysLeuHisLys 198
|||||
18900 ACATGAAACAGATATACAAAGGCTGTGAGATTAATTAAGAACTTCACA 18949
198 sasPglMetGlnGluIleLysThrPheArgLeuLysLeuGlnAsnLeuG 215
|||||
18950 GGATCAAGCTCAGAGATTAATAAATTTCAAGTTAAATTTGAAATTCGCG 18999

```

```
215 lnhThValLysaspGlnAlaHis.Lys..... 223
|||||:|||||:|||||:
19000 AGACCTGAAAGACGACCTTATAGATGTGTAAGACTTTAACTGTTTG 19049
223 ..... 223
19050 ACATGATTTTCTTTCTTTCTTCCAGTACCGACCCTGCTCTCAGCA 19099
223 ..... 223
19100 TAAATTTTTTTTGTAAAGTCAAGACAAATTAACACTAAGCTCAT 19149
223 ..... 223
19150 TGTTTTAATCTGAAAGTAATTAAGATATTGTATACATATAGTATGCA 19199
224 ..... 224
19200 TTTCTAGTGTGTGGGCTATTCTTATGCCATGATTAATCCTATACAGC 19249
224 euATgGluAsnIleAlaGlnaspGlnLysSeraspAlaSerLysSer 240
|||||:|||||:|||||:|||||:|||||:
19250 TTCTGAGAGCATTCCTCAAGATCAAGAAAGACAGAACTCTCAAAAGTT 19299
241 GlnMetGlnGlnLeuLysLysIleCysGlyThrGlnArgGluIle 257
|||||:|||||:|||||:|||||:|||||:
19300 CAGATGTGGAAGTGAACCCAGTGTTCAGAAAGTTGATGCTGAAGTTCA 19349
257 uGlnMetGluThrSerLeuaspGlnLeuArgArgLeuGlnIleAla 274
|||||:|||||:|||||:|||||:|||||:
19350 TAAATAGGAATATGATGTTAAAGATCTGAGAAAGCTCCAGAGCAAGTAT 19399
274 spIleLysAlaThrGlnArgSerThrLeuLeuThrGlnIleHisGluLys 290
|||||:|||||:|||||:|||||:|||||:
19400 CAATAAACACGCGGAAAGAGACACCTGTTTAAAGAGCAGCAGAGACAG 19449
291 LeuAlaAlaLeuSerGlnGlnAsn.Glu..... 299
|||||:|||||:|||||:
19450 TATGAGCATTAACCTGAGGAAATGAGGTTTCTTTTACTATACACAA 19499
299 ..... 299
19500 TGTTAAGTCTTCTGCTTTAGTCTGCTCCGTTTACAGATTCAG 19549
299 ..... 299
19550 TCCCTGCAATTTGTATGAATTTACTTTGCTTCTGTAAAGATCTCAC 19599
299 ..... 299
19600 ATTTGTTATTTCTTACGTAATTAATTTGTGCTGCTGTAAATATT 19649
300 ..... AspThrAspGlnGluLeuMetGluThrGlnThrLysPheGlu 313
|||||:|||||:|||||:|||||:|||||:
19650 GTTGTTCAGATATCTATGAAAGATGAAAGAAAGGAAAGCAAGTTGAA 19699
314 GlnArgIleAlaLeuLeuGlnLysIleSerLysLeuValArgAspXe 330
|||||:|||||:|||||:|||||:|||||:
19700 GAAAGACTTCCTCTTTTGGGAGCAAAATTCGAAATAAGAGAGAGAGAT 19749
330 LaspAspGlnAlaSerTyrSerSerValLeuSerLysGlnAsnSerGlu 347
|||||:|||||:|||||:|||||:|||||:
19750 GCTTGACACAGAACTACAAATTTCTGCTTCACATGCAAAAACATACT 19799
347 euThrHisGlnIleGlyLysLeuGlnAlaGluAlaAsp..... 359
|||||:|||||:|||||:|||||:|||||:
19800 ATATCTGAGATCAGCAAGCTACAAACAGAGCTGAAGTAAATCTGAAA 19849
359 ..... 359
19850 AAGCACACCATAGTAAATATATATGAGTCTCTGCTATTTGATTAA 19899

360 ..... AlaHisLeuThrMetLysHisGlu 368
|||||:|||||:|||||:
19900 CATGTTTCTTATATTTTTCATAGGCTCATATGCTGCGAGAAATGAAA 19949
368 rGAspSeraspIleLysAlaIleCysThrLysHisAsnLeuIleProVal 384
|||||:|||||:|||||:|||||:|||||:
19950 GAGATCTACATCCAGAACTATTTTCCATTAACAATTTAGAAATGTT 19999
385 ProGlnHisProPheThrAsnAspValAlaMetAsnLeuThrAsnArgTl 401
|||||:|||||:|||||:|||||:|||||:
20000 CCTAGTACCCCTTCAGTACGAGTATGTTGAAATCTCAAAATCGAAT 20049
401 eLysAlaArgLeuSerSerLeuGlnAsnAspLeuAspLys.Lys... 416
|||||:|||||:|||||:|||||:|||||:
20050 AAAATCAAGACTGGGTGAGTCTGAGATGATTTGCTCGATTAAGAAGTTC 20099
416 ..... 416
20100 ACATCTTATTTTCTACAGAAATTAATGTATATTCATTTATTTGCAT 20149
417 ..... LysSerAsnGluAla 421
20150 AATGTGACATCTTATAGAGATTTGATGTCATCTCGAAATATCAATGAAA 20199
421 spGlnLeuAspValLeuThrPlysHisTyrLeuLysIleAsnAlaArgTyr 437
|||||:|||||:|||||:|||||:|||||:
20200 CTGCTTAAGTACGCGCTGGGATGCTATATGAGATCTAAATGACGCTTG 20249
438 SerGlnValAspGlyGlnIleGlnSerLysIleGlu.Ser..... 450
|||||:|||||:|||||:|||||:|||||:
20250 AAAAGATCGAAGCTCAGAAACGAGCTAAAGATGAATCAAGGTGTGTT 20299
450 ..... 450
20300 CACCATTTGGAACCTTAACCAAAATCTCCACTAAGAGCCTAAGTTTTC 20349
451 ..... MetSerGlyIleLeuArgArg 458
20350 ACTCTTTTCTTTTATGATGCTGACTTATGCAATGAGGCAATATCAAAACGA 20399
458 rGlyAsnLysGlnLysGlnArgAspAlaAlaGluValGluLeuSerLys 474
|||||:|||||:|||||:|||||:|||||:
20400 TGAAGAGAAAGAAATGAGCGGTGATTCATTTGAAATTCGAAATTCCTACT 20449
475 PheAsnLeuSerArgTlleAspGlnArgGluArgHisMet..... 487
|||||:|||||:|||||:|||||:|||||:
20450 GTGATGTTAAACAACTGATGAAGAGAAACAAAGTGTATGCTTTC 20499
487 ..... 487
20500 GTCCCTTTATTTACTACATGATTTTGGTCTTGATGACTTTCCCTGCT 20549
488 ..... GlnIleGluVal 491
20550 AATATGCTTACTTTTTCATCATCTGCGTTTGACACAGCAAGTGAAGTTC 20599
492 GlnArgLysThrLeuAlaLeuGlnGluArgAspTyrAspSerIleIleSe 508
|||||:|||||:|||||:|||||:|||||:
20600 GAGAGAGAAACAAACAAAGCAATCTGAAAGGGATTTGAATCCAAATTAAG 20649
508 rGlnLysArgThrGlnValLysSerLeuGlnLysIleLysValLeu 525
|||||:|||||:|||||:|||||:|||||:
20650 ACAGAAACAGCATGAATATACGCTTGAGCTTAATAATCAAGACACTGA 20699
525 euATgGlnLysAspIleIleAsnArgAsnAlaAspGlnArgValLysLeu 541
|||||:|||||:|||||:|||||:|||||:
20700 ATCGAGAAAGAGCGTATGCGCTGAGAGCTGAGAGTATGAGAGAAATGG 20749
542 GlyLeuLysLysAspAlaLeuGlnLysSerLysAspLys.LysAsnGlu 557
|||||:|||||:|||||:|||||:|||||:
20750 TCTTTAAAGAAAGCGAGACAGAGATCTTAAGNAGAAACACAAAAGAT 20799
558 Ile..... 558
```

```
20800 GTGATCTCACTCACCACAACTTTCTCTGAGTTTCTGCTTTTCTAT 20849
559 .....ValasnGluHisL 563
20850 TATCATATGTTTCTACCTCTAAAGCTACTTACCAGAAATAGATGATGCA 20899
563 ysapblysiIlelyslValleuarglYarganProPhgIuLysasp 579
20900 AGGATCGAATTAGAGAGAGTCTGAAGGGGAGATTACACCCGAGAGAGAT 20949
580 MetLysLysGluIleasnGlnAla..... 587
20950 ATGAGAGAGGAAATAGTTCAGCTTGTGAGTCGCTGTTTCACTTAAGTA 20999
587 ..... 587
21000 TTATTTTGATGATTGGAATGAATCACTATTAATATAGATCTCTGA 21049
587 ..... 587
21050 TCAGCATCTCAAAACAGAGATAGCTTGTCTGTTGTTCAAGTTAT 21099
588 .....PheTyrProva 591
21100 CACATCTCAAAACATTCTTGTATTGTTTCTCTTGTGTTAGTCAAT 21149
591 lAspLysGluTyrasnGluLeuargSerLysSerGlnIuLagluing 608
21150 TGAAGGGAATATGATGACCTCAGTTGAATCCGAGAACAGACAGAAAG 21199
608 luleuLysPheThrGlnSerLysValThrAspAlaArgGluGluLeuThr 624
21200 AGGTCACATGTTGCAGATGAAGATACAGAAAGTAAACAATAGCCTATTC 21249
625 LysLeuAspArgAspMetLasp..... 631
21250 AAGCATACAGAGATACAGAAATGTAGACATGCTGTATTAATTAAGTTA 21299
631 ..... 631
21300 GATATCTATGTGCTAGTCTCTGACTAGCGCCTTAAGTTGTAATAACT 21349
631 ..... 631
21350 GCTCTTGACAAATTTTTCAGTGTACAAATATTAAGACATGTTAATA 21399
631 ..... 631
21400 ATGTGTCACTGTGTCACCCCTTTATTGATTGTTGTTCTACAAATGTAC 21449
631 ..... 631
21450 TGGATTGACATTTAAGACAGATAGGAATCATTTTCGGTGTACAGTGG 21499
631 ..... 631
21500 ATGATTAGAAAATATCTATTCACATCTGCACTGAGTACTTTATGCA 21549
632 AlAluysArgArgPheLeuAspSerLysLeuGlnSerIleLeuGlnIle 648
21550 GCAGAGAAAGATACATTGATGATCAGCTTCAGCCTTAAGACAGAAATC 21599
648 rAlaAsnValAspMetPheProLysValleuGlnAspAlaMetAsnLys 665
21600 AGTTACCATGATGCTTATACCAAAATGTTGAGTCGGCTAAGACAAATA 21649
665 rAspGluGlnLysArgLeu..... 671
21650 GAGATGACCGTAAACGTTTGTGTTAAACAGCGAGCTGTTCCATACC 21699
671 ..... 671
```

```
21700 CCACCCCTGTATGTTGGCGTAATCCAAATATATCTCTCTGCACTG 21749
672 .....Glu...AsnPhaE1 675
21750 AAAGCATAGTCTGTTCTGCTGCTGCTGTTGACCGCAATTAACATGCG 21799
675 aAsnGlyMetArgGluMetLeuAlaProPhgIuHisLeuAlaArgLys 692
21800 AAATGCTATGCGTCAGATGTTGAACCTTTGAGAAACAGCAGCTCAAG 21849
692 snHisValCysProCysCysGluArgAlaPheThrProAspGluLysP 708
21850 AGCATCTCTGCTGCTGCTGCGAGCGCTCTTTACAGCTGATGAAGAGCT 21899
709 GluPheValLysLys..... 713
21900 AGCTTATTAAGAGGTACACAAATCTCTCTGTTTGGCTTAACGAT 21949
713 ..... 713
21950 TCAATTTATTTCTGCTTCTGAATAACGGAATGCTCTTATTGAGTA 21999
714 .....GlnArgMetGlnAsnSerSerThra 722
22000 CGTAATTTCTTTCAATACAGCAAGGTAAGGCTTCAAGCAGAGG 22049
723 GluArgSerLysAlaLeuAlaMetGluSerSerAsnAlaGluAlaLeuPh 739
22050 GAGCATCTTAAGACATTAAGCAGTGAAGGCTTCAATGCTGACTGATTT 22099
739 eGlnGlnLeuAspLysLeuArgThrIleTyrAspAlaTyrValLysLeuV 756
22100 TCACACGCTAGACAAACTACAGCGCGAGTTTGAAGAAATACGCAAACTRA 22149
756 aLgluGluThrIleProLeuAlaGluLysAsnLeuAsnGlnHisLeuAla 772
22150 CTACTGAATATTATCTCTGCTGCTGAGAAACTTTGCAAGACATACAGAA 22199
773 AspGluSerGlnLysAlaGlnAlaPheAspAspLeu..... 784
22200 GAGCTGGCTCAGAACTGGAAGCTCTGATGATGAGCTTGGGATCACT 22249
784 ..... 784
22250 TGTATTTCTTGGATAGAACTGATCCAGCTGCAATTCATACGTTT 22299
784 ..... 784
22300 CTTTACAGTGAAGGTGTGCTGTTTCTGCTGCTGTTCTCTTA 22349
784 ..... 784
22350 ATCCGTGTGTCATGCACATGAGCTGGAATTTTAAATTCAGTTCATAGCA 22399
785 .....Leu.GlyValLeuAlaHis 790
22400 TCTCAATCAATATATGATGATGCTTTCAGGATTTAGGATCTCTCAGAG 22449
791 ValGlnMetAspArgAspAlaValAlaLeuLeuGlnProThrAspPh 807
22450 ATAAAGCTGACAAAGACTCGATGAGACACTGAGTGCAGCCCTCGAATA 22499
807 rIleAspArgHisValHisGluIleGlnGlnLeuValLysGluValGlu 824
22500 TGCTACACAGGATCTTTCAGGAATGCTATCCACAAAGCAAAATGAG 22549
824 sPheGluTyrAlaLeuAspSerSerGlyArgGlyValLysSerLeuGlu 840
22550 ATCTTGAATATAGCTTATTTAGGGGCTTGGGCTTAAACATATGCAA 22599
841 GluIleGlnLeuGluLeuAsnPhaLeuGlnArgThr..... 852
22600 GAAATCAATCAATGAATTAAGTACCTCCAAAGCAGCAGTAGTACTGTG 22649
```

852 852
22650 AATTATGACAGGTAGAACATATTTCTCTCTGAGCTTCTTAGTCT 22699
853 Arg 853
22700 AGCAGACACATATTTCCAAATTTTCTCTCTTAGTACCTTGACAGG 22749
854 AsprhLeuIleValIgluValAspAspLeuArgAspGlnHisArgMetLe 870
22750 GATAAATTGCACGCTGAGCTGGAGAACTTAGAGATGACCATTTACT 22799
870 uasnIuAspMetSerSerAlaGlnValArgTyrHisAsnAlaArgGlu 887
22800 GGAACGTATATATCATGCTCCACAGCCAGATGCGATCGCATGAGGAGG 22849
887 LuIyValIyLysAlaSerSerIleLeuGluArgPheGlnIyLysSerGlu 903
22850 AGAAGGCCAGGACGCAATTTGTTACGGGAGCTTACAAAGCGAGAA 22899
904 GluLeuValIleuLeuAlaGluGluIyLysGluGlnLeuIleValIgluLys 920
22900 GACCTGGAGCCGTTAGCCGAGGAGAAAGTCACTGACCTGATGTGA 22949
920 sLeu..LeuGlu..... 923
22950 GGTGGGTTGATGAGCGTTTATGATTGTTTCCACATGCTTTCGTGTG 22999
923 923
23000 TCACATATCAAGTCCATTGTTAATGGGAGCTATAAAGCTCTAATT 23049
923 923
23050 ATTAGTCTATGCTTAAATAGAGACTACTATGTGAAATGTGCA 23099
923 923
23100 TATCAAACTCCATTTTTCATGTGACTATAAAGTCTAATTATTAGG 23149
923 923
23150 TCCTTGATCATGAATGAGAGTACTACTTGTGAGAAATGCTAAACAG 23199
923 923
23200 ATTAGTAATCATTTTCTTCTGAATAAATTGCTTACTAGTAATGCAAT 23249
923 923
23250 GCATGCAACAGACATATATAGAAATACATAATAAATGGTAGCAATA 23299
923 923
23300 AGTACATCTCAGAGGTGATGATTTTCTGTGAATGCTTTGAGTT 23349
923 923
23350 TTTGTATTGAAGCTTCAAAATATATTTTCTGTGCAATTCGTTGAATG 23399
923 923
23400 TGATCAAGTACTCTTTCGAACATGCAAAAGTAATATCAGTCT 23449
923 923
23450 CTTCTATTTTGTTCAAATTTCTCTGTACATTAATGAGATCT 23499
924 GluSer 925
23500 TTAATAATTAATGTTCATTTCAATATTTTTCAGTATTTGACCGAGGCC 23549

926 LeuAAspProLeuSerLysGluLysGluSerLeuLeuGlnGluTyrAsnAl 942
23550 CTGGCCCTTTTCCAGAGAAAAGCAATTAATCTGAGTACTAAGTGA 23599
942 sLeuLysGlnLysLeuAspGluGluTyrHisGlnLeuAlaGluArgLys 959
23600 TATGAAGATTACAGTAATCAAGAGTATGAGAACTGCTGAGAAAAA 23649
959 rGluPheGlnGlnLysLeuAspAlaLeuGlyArgLeuAsnMetLysIle 975
23650 GAAACTACCAACAGAGAGTGTAGCATTAATCAAGCCAGTATTAATA 23699
976 LysGlyTyr..... 978
23700 AATGAGTATGTTCTGCTAGCAAGGTTAATTCATTTATCTCTGCAAC 23749
979 Leu..AspSerLysLysAsnGlu 985
23750 AGATATCCTTACAGCTGCTTACAGGATTCATGATTCMAAGAGGGA 23799
985 uLysLeuLysGluLeuGlnGlyArgHisValLeuLysHisSerGlnLeu 1002
23800 GCGCTTAGTGAATTCAGAGAAAGCAGGCTTATCTGATTCAGCTTC 23849
1002 LysSerCysMetAlaLysGlnGlnArgIleSerAlaGluLeuAsnLysSer 1018
23850 AGAGCTGTAGAGCTAAGAAAGAAATGAATCTGACAGTGAAGCAAAAT 23899
1019 LysGluLeuLeuGlnGlnGlyGlnGlnLysLeuLysArgAsnIleAspAspAs 1035
23900 AAAGCTTATGCGGAAATCAAGATCAATGAGAAATATAGAGATTA 23949
1035 nLeuLysTyrArgLysThrLysAlaAspValGluGlnLeuThrArgAspI 1052
23950 CTTGAACCTACCGGACAAAGCAAAAGTTGAGGAGCTTACACGTGAGA 23999
1052 LeuLysSerLeuGluGluArgLeuLeuSerIleGlySerLeuSerAlaIle 1068
24000 TTGAATCATTTGGAAGAGCAAAATTTGAAATTTGGTGGAGTTGCCAGT 24049
1069 GluAlaAspLeuLysArgHisSerGlnGluLysGluArgLeuAsnSerG 1085
24050 GAAGCTGAATTTGTAAAGATTTGCGGAAAGGAAAGCTTCTTCA 24099
1085 u..... 1085
24100 GGTAAACAGATCCTTATATATTTATAGCATCATCATGATATGATTT 24149
1085 1085
24150 GGTGTATAGTTAGCTATTCATTTGTTCTCCATGATGCTAAGTAA 24199
1086 PheAsn..Arg..Tyr..... 1089
24200 TCTTGATTTCTCTGATTTAAGCATTTAATTCGAGAGTGTGTTT 24249
1089 1089
24250 AAAATTAATCCTGTGTTCAAGTAGCCAAATATATAAATTTAGAAA 24299
1089 1089
24300 TTATCATTTATAGATAACAGATTTCTCTACGTACATTCATTCGCTG 24349
1089 1089
24350 CTGTTACAACTGATATGTGTGAACGTGAGTTAAGTCTGTTTCTGA 24399
1089 1089
24400 TTATTTAAGATAGTTATCATTTGAAGGTTTCAGAGCTATTTATTTG 24449
1090 GlnGlyT 1092


```

gene
    /cell_line="HeLa"
    1..449
    /gene="RAD50"
    /note="DNA recombination repair gene"
    322..4260
    /gene="RAD50"
    /function="DNA recombination repair protein"
    /note="splice variant: RAD50-2 has a shorter 3'-UTR than
    the previously reported RAD50 transcript encoded by the
    sequence presented in Genbank Accession Number U61339"
    /codon_start=1
    /product="RAD50-2 protein"
    /protein_id="RAD50325.1"
    /db_xref="GI:5739041"
    /translation="MSRIEKMSTLVGRSPCEIDKDKIITPESPRTLVGNAGKWT
    IIECLVITGDPFGKGTGTVHDPKVAOETRADIRLQFQVNSGLAVQSSNYC
    TQSKTEFTLEGVITRTKHGKSVLSKCAEDRMISLGVSKAVLNNVJCHQE
    DSNWPLSEKALOKFDEISATRYIKALETLFOVOTOCQKVEYOMELKLYOE
    KACEIDQITSKEAOLTSKEIVKSYENELDPKRLKEIHNLSKIMKLDNEKALD
    SRKKOMEKNSLEIEKMEKVFQGTDEQINDLNHQBTHREKELVDCHELEKMK
    ESRLLNOKSELVEQGRLOLQADROEHTRARDLSLOSLATQELGREGPSESEK
    IKNHMLVPEROGEAKTAQNLANDPFAKETLQKQIDELDKKTGLGRILIELSEL
    SKQNELKNVYELQQLQEGSSDRILELQELIKAERLSKAKNSVETLMEYISIQ
    NEKADRLRTKLDQEMOENLHHTTTRQKEMLTDKYKADQEQRIKIKSRHDELTL
    LGYPMNKOLEMLHKSKEINOTROBLAKNELASQNKHINNELKREKOLSS
    YEDKLDVQCSQDPESDLDLKEIEKSSKQRLAKATAVYSQFIQLDVENOSCP
    VQVRFQTEALELQEVISDLSKRLADPKLSTESLKKERKEDMLGLVPMOSITL
    DLKEKEIPELRNKLQNVNRDIOQLKNDIEQDETLLGTIMPEESAKVCLDVTIMER
    OMELKVERKIAQQAALQGLIDRLTVQVQOEKQKIDTVSSKIELNRKLIQDQ
    QEOIHLKSTTNELSEKLIQISTNLROOLEEOTVELSTVQSLYEIDAKQVAP
    LETTLEKFOEKELIKKTKNSKINQDKNDIEKKNHIGMKQIDENYIQDCKDY
    KKOETEELNKVIAQLSECEKHEKINEDMLMRDITQKQERNLDNLTREKNEE
    LKEVEERKQHLKEMQOVLOKMSQKLEENIDNLRNNLLHGRKHSMEBINIIF
    KKELEKREPOFDAEKYREMIYMKETELVKNKLDIDYKTLQAKHSMKEBINI
    IRDLARSTYEGODIEYIEISRDADENASASDKRNNYRVYMLLGLVLEIKRSQCSG
    OKYLASTIIRLALAEFLKNGIATLDEPTNDRNENESLHVLDELKISRSQGRNF
    QLVYITDEPVELIGRSEYVERKIRKKNIDQSEIVKCVSSLGKRVH"

BASE COUNT      1726 a      753 c      978 g      990 t
ORIGIN

alignment_scores:
    Quality: 1730.00      Length: 1385
    Ratio: 1.918          Gaps: 38
    Percent Similarity: 65.126      Percent Identity: 32.419

alignment_block:
    US-09-538-396-2 x AF057299 ..

Align seg 1/1 to: AF057299 from: 1 to: 4449

```

```

569 GTCGCAATTTCGTGATGTCATGAGACCTTATAGCTGTGCAAGATCT 618
99 PheGluLeuThrGlnIleValAlaSerIleMetGluPheLysAlaIleGlu 115
619 ATGTGTGTACTCAGAAAGACAAAGACAGATTTAAACTCTGGAAGG 668
115 rValLeuGlnThrIleAsnProHisThrGlyIleValCysLeuSer 132
669 AGTCATT..ACTAGAACAAAGCAT...GGTAAAAAGGTCAGTGTGACT 712
132 yTArgCysAlaAspMetAspArgGluIleProAlaLeuMetGlyValSer 148
713 CTAAAGTGTGCANAAATGACCGAGAAATGATCAGTCTCTGGGGTTCC 762
149 LysAlaValLeuGluAsnValIlePheValHisGlnAspGluSerAspTr 165
763 AAGGTGTGCTTAATTAATGATCTTTCTGTCTCAAGAAATTCCTAATG 812
165 ProLeuGlnAspProSerThrLeuLysLysLeuAspAspIlePheS 182
813 GCCTTAAAGTGAAGAAAGCCTTGAAGCAAAAGTTGATGATTTT 862
182 eValAspThrArgTyrThrIleValAlaLeuGluValIleLysLysLeuHisLys 198
863 CAGCAACAAGATACATTAAAGCCTTAGAAMACCTTCGAGGTAGCTGAG 912
199 AspGlnMetGlnGluIleLysThrPheArgLeuLysLeuGlnAsnLeu 215
913 ACACAAAGTCAGAAAGTAAAGATATCAAGAAATGCAAAATATCTGAA 962
215 nThrValLysAspGlnAlaHisLysLeuArgGluAsnIleAlaGlnAsp 232
963 GCATATTAAGGAAAAAGCTTGAGATTTGCTGATCAGATTAACAATAG 1012
232 LngLysSerAspAlaSerLysSerGlnMetGlnLeuLysGluLys 248
1013 AAGCCAGTAACTCTCAAGAAATGTCAAAATCTATAGATGAA 1062
249 IleCysGlyThrGluArgGluIleLeuGlnMetGluThrSerLeuAsp 265
1063 CTGTGATCATTGAAGAAATGCTTAAAGAAATGAAATGAAATCTCTTAA 1112
265 uLeuArgArgLeuGlnGlyGlnIleAspIleLysAlaThrGluArgSer 282
1113 AATTAATGAACCTTGACATGAATTTAAACCTTGATACCCAGAAAGAC 1162
282 hLeuLeuThrGlnGlnHisGluLysLeuAlaIleLeuSerGluGluAsn 298
1163 AATGAGAAAGATATATAGTGAAGTGAAGAAATGAAAGGTTT 1212
299 GluAspThrAspGluLeuMetGluTTPGlnThrLysPheGluGluArg 315
1213 CAAGGACGTAGTGAACAACTAAATGACTTATATCAACAAACACCAAGAAC 1262
315 gIleAlaIleLeuGlnThrLysIleSerLysLeuValArgAspMetAsp 332
1263 AGTAAGGAGAAAGAAAGAAATGCTAGACTCTCATCTGAACTGAAA 1312
332 sPGLuAlaSerLysSerSerValLeuSerLysGlnAsnSerGluLeuThr 348
1313 AACTAATAAAGAAATCTAGGCTTCTCAATCAGAAAATCAGAACTGCT 1362
349 HisGluIleGlyLysLeuGlnAlaGluAlaAspAlaHisLeuThrMet 365
1363 GTTGAACAGGCTGCTCTACAGCTGCAAGCAAGATCGCCATCAGAACATAT 1412
365 sHisGluArgAspSerAspIleLysAsnIleCysThrLysHisAsnLeu 382
1413 CCGAGCTAGAGATTCATTAATTCAGTCTTGGCAACACAGCTGAATTTGG 1462
382 LyrProValProGluHisProPheThrAsnAspValAlaMetAsnLeuThr 398
1463 ATGCGTTTGAAGCGTGAACCATTCAGTGAAGAAAGACAGATTAATTTTCA 1512

```


399 AsnArgIleuAlaArg.....LeuSerSerIe 408
1513 AAACCTGTGTAGAGACAGCAAGGGAACCAAAATCTCCAAACCACT 1562
408 UGLAspLeuLeuAspIlySylSerAsnGluAspGluLeuAspV 425
1563 GATGATGACTTTCAGAAAAGAGACTCGAAAACAAAACAGTAAAT. 1611
425 alleuTrpLysHisTyrLeuLysIleAsnAlaArgTyrSerGluValAsp 441
1512GAGTAAAGATTAAGAAAACAGCTGCGGA 1641
442 GlyIleIleGlnSerLysIleGluSerMetSerGlyIleLeuArgArg 458
1642 AGAATAATGTGTTAAATCAGAAATCTAACT.....AA 1676
458 GLYAspLysGluLysGluArgAspAlaAlaGluValGluLeuSerLysP 475
1677 GAGCAGAAATGAGCTGAAAATGTGAAGTATGAATTCACAGACTGGAG 1726
475 heaenLeuSerArgIleAspGluArgGluArgHisMetGlnIleGluVal 491
1727 GATCTCAGACAGGATCTGGAAGTACAGCAGAGCTC...ATGAAGCT 1773
492 GluArgLysThrLeuAlaLeuGlyGluArgAspTyrAspSerIleIleSe 508
1774 GACGTGAG.....TTAAGCAGGCTGAGAAAACAGCAATGTAGA 1814
508 RGLuLysArgThrGluValTyrSerLeuGluLys..... 520
1815 AACCTTAAATAAGAGATTAATAGTCTCCAAATGAAAAGCAGACTTAG 1864
521IleLysValLeuLeuArgGluLysAspIleIleAsnArgAsn 534
1865 ACAGAGACCTCTGTAACCTGACAGAGAGAGAGAGCTTAACATCAT 1914
535 AlaAspGluArgValLysLeuGly...LeuLysLysAspAlaLeuGluSe 550
1915 ACAACAACACCTACCCAAATGAGATGCTGACCAAGACAAAGCTGACAA 1964
550 rSerLysAspLysLeuAsnGluIleValAsnGluHisLysAspLysIle 567
1965 AGAT...GACAAATTCGAAAATTAATCTAGGCACAGTATCAATTA 2011
567 yLysValLeuArgLysArgAsnProPheGluLysAspMetLysLysGlu 583
2012 CCTCAGCTTG...GGATATTTCCCAACAAAACAGCTGAGAGC... 2055
584 IleAsnGluAlaPheTrpProValAspLysGluTyrAsnGluLeuArgSe 600
2056TCG.....CTGCAATAG 2066
600 rLysSerGlnGluAlaGluGlnGluLysPheThrGlnSerLysValT 617
2067 TAAATCAAAAAGAAATTAATCAG..... 2088
617 hrAspAlaArgGluGlnLeuThrLysLeuArgAspMet..... 630
2089ACGAGGACAGACTTCCAAAATGAGCAAGAGACTGCTCATCT 2133
631 AspAlaLysArgArgPheLeuAspSerLysLeu...GlnSerIleLeuG 646
2134 GAGCAGAAATTAATCAATTAATTAATCACTAAAAGAAAAGAGAGACA 2183
646 nIleSerAlaAsnValAspMetPheProLysValLeu.....GlnAspA 661
2184 GTTGTCCAGTTACGAGACAGAGCTGTTGATGTTGTGTGAGCCAGGAT 2233
661 IametaLysLysArgAspGluLysArgLeuGluAsnPheAlaAsnGly 677
2234 TTGAAGAGTATTAGACAGGCTTAAAGAGAAATGAAAATCATCAAAA 2283

678 MetArgGluMetLeuAlaPro.....PheG 686
2284 CAGGAGCCATGTTGCTGGAGCCACAGACTTACTCCAGTTCATTA 2333
686 uHisLeuAlaArgLysAsnHis...ValLysProCysCysGluArgAlaP 702
2334 TCAGCTAACAGAGAAAACCATGATGTTGCCCCGTTGTGACAGAGATT 2383
702 he...ThrProAspGluGluAspGluPheValLysLysGlnArgMetGln 717
2384 TTGAGACAGAGGCTGAGTACAAAGATCATCACTGATTGTCAGCTTAA 2433
718 AsnSerSerThrAlaGluArgSerLysAlaLeuAlaMetGluSerSer 734
2434 CTGCGACTTGTCTCAGATTAATCACTACAGACAGAAATCAAGCTTAAA 2483
734 nAlaGluAlaLeuPheGlnLeuAspLysLysLeuArgThrIleTyrAsp 751
2484 AAGGAAAAGGCGGCTGATGAATGCTGGAGCTGTGCCATGAGGCAA 2533
751 IatYValLysLeuValGluGluThrIlePro...LeuAlaGluLys... 765
2534 GCATTAATTTGATTTCAGAGAGAGAAATACCGAATTAAAGAAAACCTG 2583
766 ...AsnLeuAsnGlnHisLeuAlaAspGluSerGlnLysAlaGlnAlaP 781
2584 CAGATGTCAATAGACATACAGCCCTAAGAACATAGAACAGACA 2633
781 eAspAspLeuLeuGly..... 786
2634 AGAACAACCTGTGGTACAAATATGCTGGAAGAGAAAGTCCAAAGTAT 2683
787ValLeuAlaHisValGlnMetAspArgAspAla 797
2684 GCCTGACAGATGTACATTAATGAGAGAGTTCCAGATGGAATTAAGAT 2733
798 ValGlu.....AlaLeuLeuGlnProThrAspTh 807
2734 GTTGAAGAAAATTCGACAAACAGCACTAGCTACAGAAATAGAC... 2781
807 rIleAspArgHisValHisGluIleGlnGlnLeuValLysGluValGlu 824
2782 .TTAGATCGAAGCTGCCAACAGTCAACCAAG.....GAGAAACAG 2821
824 sPLeuGluTyrAlaLeuAspSerSerGlyArgGlyValLysSerLeuGlu 840
2822 AGAACAAGCAACAGTACAGACA.....GTTCTAGTAAAGATT 2859
841 GluIleGlnLeuGluLeuAsnPheLeuGlnArgThrArgAspThrLeuI 857
2860 GAATTAATGCTAGCTT..... 2877
857 eValGluValAspAspLeuArgAspGlnHisArgMetLeuAsnGluAsp 874
2878ATACAGACAGCAGAGAACAGATCAACATCTAA 2911
874 eLysSerAlaGlnValArgTrpHisAsnAlaArgGluGluLysValLys 890
2912 AAATGACACA.....AATGACTTAAATCTGAAATCACTTCCAG 2949
891 AlaSerSerIleLeuGluArgPheGlnLysSerGlnGluGluValLe 907
2950 ATATCCACTAATTTGCAACGCTGTCAGCAACTGGAGGAGCACTGTGGA 2999
907 uLeuAlaGluGluLysGluGlnLeuIleValGluLysLysLeuLeuG 924
3000 ATTATCCACTAAGTCAAGTCTTTGTACAGAGAGATTAAGAGATCTAAG 3049
924 IuSerLeuAspProLeu.....SerLysGluLys 933
3050 AGCAGGTAAAGCCCTTGAACAAACATTTGAAAAGTTCCACCAAGAAAA 3099
934 GluSerLeuLeuGlnGluTyrAsnAlaLeuLysGlnLysLeuAspGluG 950


```

3390 TGAGCTCATGACAGATATGTGATACAGAGATACAGAGAGAGTGG 3439
1032 ILeaspaspaenleuLysTyrArgLysThrLysAlaAspValGluGlnLe 1048
3440 CTACAGATTAACCTTACTTTTAAAGAAAGATGAGAACTAAAGAAAGT 3489
1048 uThrArgspIleGluSerLeuGluArgLeuLeuSerIleGlySerL 1065
3490 TGA.....GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3530
1065 euserAlaIleGluAlaAspLeuLysArgHisSerGlnLys..... 1079
3531 TGCAGGCTTTCAGATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3580
1080 GluArgLeuAsnSerGluPheAsnArgTrpGlnLysThrLeuSerValTy 1096
3581 GACATATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3630
1096 rGlnSerAsnIleSerLysHisLysGlnLysLeuLysLeuSerGlnTyrL 1113
3631 TGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3680
1113 yAspIleGluLysArgTyrThrAsnGlnPheLeuGlnLeuLysThrThr 1129
3681 GCGATGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3730
1130 GluMetAlaAsnLysAspLeuAspArgTyrTyrThrAlaLeuAspLysAl 1146
3731 GAACCTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3780
1146 AleuMetArgPheHisSerMetLysMetGluGlnLeuAsnLysIleLeu 1163
3781 AATATGTAATTTACAGATAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3830
1163 ySgLeuTrpGlnGlnThrTyrArgGlyGlnAspIleAspTyrIleSer 1179
3831 GTGACCTGTGGCAGAGTACATCTGTGAGAGAGAGAGAGAGAGAGAG 3880
1180 IleAsnSerAsp.....SerGlnLysAlaGlyThrArgSe 1191
3881 ATACGCTGTGATGCCGATGAAATGTATCAGCTCTGTATTAAGCGGGA 3930
1191 rTyrSerTyrArgValValMetGlnThrGlyAspAlaGluLeuGluMet 1208
3931 TTATATACACGAGTGTGATGCTGAGAGAGAGAGAGAGAGAGAGAG 3980
1208 rGlyArgCysSerAlaGlyGlnLysValIleuAlaSerLeuIleIleArg 1224
3981 GAGGACGATGCACTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4030
1225 LeuAlaLeuAlaGluThrPheCysLeuAsnGlyIleLeuAlaLeuAs 1241
4031 CTGGCCCTGGCGGAGAGCTGTGCTCAACTGTGAGATTCCTTGA 4080
1241 pGluProThrThrAsnLeuAspGlyProAsnAlaGluSerLeuAlaAla 1258
4081 TGAGCCACACAAATCTGACCGAGAGAGAGAGAGAGAGAGAGAGAG 4130
1258 laLeuLeuArgIleMetGluAlaArgLysGlyGlnLysAsnPheGlnLeu 1274
4131 CTCGTGTGAGATATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4180
1275 IleValIleThrHisAspGluArgPheAlaHisLeuIleGlyGlnArgL 1291
4181 CTGGATATCAGATGAGATATTTGTGGAGCTTTAGGAGCTTGA 4230
1291 nLeuAlaGluLysTyrTyrArgValSerLysAspGluAsnGlnHisSer 1308
4231 AATGTGTGAGAGAGATTTCTACAGATTAAGAGAGAGAGAGAGAGAG 4280
1308 leIle 1309

```

```

4231 AGATT 4285
seq_name: gb_pat1:AR079256
seq.documentatlon_block: 5893 bp DNA PAT 31-AUG-2000
LOCUS AR079256
DEFINITION Sequence 44 from patent US 5965427.
ACCESSION AR079256
VERSION AR079256.1 GI:10006002
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 5893)
AUTHORS Dolganov, G. and Novikov, A.
TITLE Human RAD50 gene and methods of use thereof
JOURNAL Patent: US 5965427-A 44 12-OCT-1999;
FEATURES
source Location/Qualifiers
BASE COUNT 2081 a 1097 c 1301 g 1414 t
ORIGIN
alignment_scores:
Quality: 1730.00 Length: 1385
Ratio: 1.918 Gaps: 38
Percent Similarity: 65.126 Percent Identity: 32.419
alignment_block:
US-09-538-396-2 x AR079256
Align seg 1/1 to: AR079256 from: 1 to: 5893
1 MetSerThrValAspLysMetLeuIleLysGlyIleArgSerPhe.... 15
389 ATGCCCGGATCGAAGAGATGAGCATCTGGCGGTGCGAGATTTGGAT 438
16 AspProAsnLysAsnValIleThrPhePheLysProLeuThrLeuI 32
439 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 488
32 leValIleProAsnGlyAlaGlyLysThrThrIleIleGlyCysLeuLys 48
489 TGGTTGAGCCCAATGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 538
49 LeuSerCysThrGlyGluLeuProAsnSerArgSerGlyHisThrPh 65
539 TATATTTCTAGGAGATTTCCCTCTGAGAGAGAGAGAGAGAGAGAG 585
65 eValHisAspProLysValAlaGlyGluThrGluThrLysGlyGlnIle 82
586 TGTACACATCCCAAGGTGCTCAAGAGAGAGAGAGAGAGAGAGAGAG 635
82 yLeuArgPheLysThrAlaAlaGlyLysAspValValCysIleArgSer 98
636 CTCCTCAATTCGTGATCTCAATGAGAGAGAGAGAGAGAGAGAGAGAG 685
99 pHeGlnLeuThrGlnLysAlaSerLysMetGluPheLysAlaIleGluSe 115
686 ATGGGTGTACTACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 735
115 rValLeuGlnThrIleAsnProHisThrGlyGluLysValCysLeuSer 132
736 AGTCATT...ACTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 779
132 yArgCysAlaAspMetAspArgGluIleProAlaLeuMetGlyValSer 148
780 CTAAGTGTGCAAGATTAACGAGAGAGAGAGAGAGAGAGAGAGAGAG 829
149 LysAlaValIleGluAsnValIlePheValHisGlnAspGluSerAsnTr 165
830 AAGCTGTGCTAATATATGATTTCTGTGATCAAGAGAGAGAGAGAGAG 879

```

165 pProLeuGlnAspProSerThrLeuLysLysPheAspIlePheS 182
182 eAlaIleThrArgThrLysAlaLeuGlnValIleLysLysLeuHisLys 198
930 CAGCAACAAGATACATTAAAGCCTTAGAACAACCTCGCAGCTAGCTGAG 979
199 AspGlnMetGlnGluIleLysThrPheArgLeuLysLeuGlnAspGlu 215
980 ACACAAGCTGAGAAAGTAAAGAAATGCAATGGAATGAAATATCTGAA 1029
215 nThyValLysAspGlnAlaHisLysLeuArgLysAsnIleAlaGlnAspG 232
1030 GCAATATATAGGAAAAAGCTGTGAGATTCGTATCAGATTACAACTAAG 1079
222 GlnLysSerAspAlaSerLysSerGlnMetGlnLysGlnLysGlnLys 248
1080 AAGCCCACTTAACATCTTCAAGGAAATGTCAAATCTCATGACAAATGA 1129
249 IleCysGlyThrGluArgGluIleLeuGlnMetGluThrSerLeuAspG 265
1130 CTGTGATCATGAGAAATCGCTTAAGAAATGAAATCAATCTCTCTAA 1179
265 uLeuArgArgLeuGlnGlnIleAspIleLysAlaThrGluArgSerT 282
1180 AATATGAACTTGACAAATGAAATTAAGCCTTGATGATGCCGAAAGAAC 1229
282 hIleLeuThrGlnGlnHisGlnLysLeuAlaIleLeuSerGlnLysAsn 298
1230 AAATGAGAAAGATATATAGTGAAGTGAAGAAAGAAATGAAAGGTTT 1279
299 GluAspThrAspGluGluLeuMetGluThrGlnThrLysPheGlnGluArg 315
1280 CAAGGACTGATGACCAACTAAATACCTTATTCACAAATCACCAGAGAAC 1329
315 gIleAlaLeuLeuGlnGluThrLysIleSerLysLeuValArgAspMetAsp 332
1330 AGTAAAGGAGAAAGAAAGAAATGCTGACACTGCTCATCTGAACTGAA 1379
332 spGluAlaSerTyrSerSerValLeuSerLysGlnAsnSerGluLeuThr 348
1380 AACTAAATAAAGCAATCTAGCTTCATCAATCAGGAAAAATCAGACGCTT 1429
349 HisGluIleGlyLysLeuGlnAlaGlnAlaAspAlaHisLeuThrMetLys 365
1430 GTTGAACAGGCTCGCTACAGCTGCAAGCAGATCCCATCAACAACATAT 1479
365 sHisGluArgAspSerAspIleLysAsnIleCysThrLysHisAsnLeuG 382
1480 CCGAGCTAGAGATTTCATTAATCAGCTTTGGCAACACAGCTAGAATTGG 1529
382 LysProValProGluHisProPheThrAsnAspValAlaMetAsnLeuThr 398
1530 ATGGCTTGAGCGCTGACCATTCAGTGAAGACAGATTAAATAATTTTCA 1579
399 AsnArgGlyLeuAlaArg.....LeuSerSerLeu 408
1580 AAACCTTGAGAGAGACAGACAAAGGCAAGCAAAACTGCCAACCAACT 1629
408 uGluAsnAspLeuLeuAspLysLysSerAsnGlnAspGlnLeuAspVal 425
1630 GATGATAGTACTTGCAGAAAAAGACACTGAAACAAACAAACAGATGAT 1678
425 aLeuThrLysHisTyrLeuLysIleAsnAlaArgTyrSerGluValAsp 441
1679GAGATAGAGATAAAGAAACTGCACCTGGGA 1708
442 GlyGlnIleGlnSerLysIleGlnSerMetSerGlyIleLeuArgArgArg 458
1709 ACAATATATTGAGCTTAATCAGAAATCTTAAGT.....AA 1743

458 gLysAspLysGlnLysGluArgAspAlaIleGlnValGlnLeuSerLysP 475
1744 GAGCAGAAATGAGCTGAGAAAAATGTGAAGTATGAATATACAGAGTTGAA 1793
475 hAsnLeuSerArgIleAspGluArgGlnGlnHisMetGlnIleGlnVal 491
1794 GATCTTCAGACGAGATCTTGAACCTGACACGAGACTC...ATAAACT 1840
492 GluArgLysThrLeuAlaLeuGlnGluLysArgAspTyrAspSerIleIle 508
1841 GAACTGTAG.....TTAAGCAAGCTGAGAAAAACAGCATGTAGA 1881
508 rGlnLysArgThrGlnValTyrSerLeuGlnLys..... 520
1882 AACCTTAATAAGAGAAATATAGTCTCCAAATGAAATGAAACAGACTGAG 1931
521IleLysValLeuLeuArgGluLysAspIleIleAsnArgAsn 534
1932 ACAGAGCCCTGCGTAACCTTGACAGAGATGACAGCACTTAACCATCAT 1981
535 AlaAspGluArgValLysLeuGly...LeuLysLysAspAlaLeuGluSe 550
1982 ACMAACACGCTACCCAAATGAGATGCTGACCAAGACAAAGCTGACAA 2031
550 rSerLysAspLysLeuAsnGluIleValAsnGlnHisLysAspLysIleL 567
2032 AAT...GACAAATCAGAAAAATTAATCTAGCAGACAGTGTGAATTA 2078
567 yAlaValLeuArgGlyArgAsnProPheGlnLysAspMetLysLysGlu 583
2079 CCTCACTGTG...GGATATTTCCCAACAAACAAACAGCTTGAAAGC... 2122
584 IleAsnGlnAlaPheThrProValAspLysGluTyrAsnGlnLeuArgSe 600
2123TGG.....CTACATAG 2133
600 rLysSerGlnGlnAlaGlnGlnLysLeuLysPheThrGlnSerLysValT 617
2134 TAAATCAAAAGAAATTAATCAG..... 2155
617 hAspAlaArgGlnGlnLeuThrLysLeuArgArgAspMet..... 630
2156ACGAGGACAGACTTGCCAAATGAAACAGAACTCACTCATCT 2200
631 AspAlaLysArgArgPheLeuAspSerLysLeu...GlnSerIleLeuGln 646
2201 GAGCAAAATAAATCATATTAATATGAACTAAAGAAAGAAAGAGAGA 2250
646 nIleSerAlaAsnValAspMetPheProLysValLeu.....GlnAsp 661
2251 GTTGTCCAGTTACGAAGACAGACTGTTGATGTTGTGTGTACCCAGAT 2300
661 lMetAsnLysArgAspGlnGlnLysArgLeuGlnAsnPheAlaAsnGly 677
2301 TTGAAGTGAATTAAGCAGCTTAAGAGAAATGAAAAACATCAACAA 2350
678 MetArgGlnMetLeuAlaPro.....PheGln 686
2351 CAGCGAGCATGCTGCTGAGACACAGAGTTACTCCAGTTCTATAC 2400
686 uHisLeuAlaArgLysAsnHis...ValCysProCysGlnLysArgAlaP 702
2401 TCAGCTTAAGAGAGAAACAGCATGTTGCCCCCTTTGTTCAGAGAGTT 2450
702 hE...ThrProAspGlnGlnLysAspGluPheValLysLysGlnArgMetGln 717
2451 TTCAGACAGAGCTGAGTTCACAAAGAGTCAATGATATTGGACGTGTAA 2500
718 AsnSerSerThrAlaGluArgSerLysAlaLeuAlaMetGlnSerSerS 734
2501 CTCGCACTTCTCCAGATTAACATCAAGTCAACAGATCAGAGCTAATAA 2550
734 nAlaGlnAlaLeuPheGlnGlnLeuAspLysLeuArgThrIleTyrAspA 751

```

2551 AAAGGAAAAAGCCGCTGATGAAATGCTGGACTGTGCCCCATGAGCGAAA 2600
751 laTyValylLeuValGluGluThrIlePro...LeuAlaGluLys... 765
::: ||| |||::: |||::: |||::: |||
2601 GCATTAATGATTGAGAGAAAGAAATACCAAGATTACAGAAACAACTG 2650
766 ...AbnLeuasnGlnHisLeuAlaAspGluSerGlnLysAlaGlnAlaPhe 781
::: |||::: |||::: |||::: |||
2651 CAGATGTCAATAGACATACAGCGCTTAAAGAACGACATAGAAAGACA 2700
781 eaSPaSPLeuLeuGly... 786
::: |||::: |||::: |||::: |||
2701 AGAAACACTTGGGTGACATTAATGCTGAAAGAAAGAACTGCCAAAGAT 2750
787 .....ValLeuAlaHisValGlnMetLaspArgAspAla 797
::: |||::: |||::: |||::: |||
2751 GCCTGACAGATGTACATATATGAGAGGTTCCAGATGGAACATAAGAT 2800
798 ValGlu... 807
::: |||::: |||::: |||::: |||
2801 GTTGAAAGAAAAAATTGCACAAAGACGACTAGCTACAAAGGAATAGAC.. 2848
807 rILeaSPArgHisValHisGluIleGlnIleValLysGluValAlaGlu 824
::: |||::: |||::: |||::: |||
2849 ...TTAGATGAACTGTCCAAAGTCAACACAG...GAGAAACAG 2888
824 sPLeuGluTyAlaLeuAspSerSerGlyArgGlyValLysSerLeuGlu 840
::: |||::: |||::: |||::: |||
2889 AGAAACAGCACAAGTTAGACACA...GTTTCTAGTAAGAT 2926
841 GluIleGlnLeuGluLeuAsnPheLeuGlnArgThrArgAspThrLeuI 857
::: |||::: |||::: |||::: |||
2927 GAATGTAATCGTAGCTT... 2944
857 eValGluValLaspArgLeuArgAsnGlnHisArgMetLeuAsnGluSPm 874
2945 .....ATACAGACACGACGAGAACACGATTCACACATCTTAA 2978
874 eTSPSerAlaGlnValArgThrPheHisAsnAlaArgGluGluLysValLys 890
::: |||::: |||::: |||::: |||
2979 AAAGTACACACA...AATGAGCTAAATATCGAAGAACTTCAG 3016
891 AlaSerSerIleLeuGluArgPheGlnLysSerGluGluGluLeuValLe 907
::: |||::: |||::: |||::: |||
3017 ATATCCACTAATTGTCACAGCTGTCACAGCACTGAGGAGCACTGTGGA 3066
907 uLeuAlaGluGluLysGluGlnIleValGlnLysLysLeuLeuGluG 924
::: |||::: |||::: |||::: |||
3067 ATTATCCACTGAAGTCACTCTTGTACAGAGATTAAGGATGCTTAAAG 3116
924 lUserLeuAspProLeu...SerLysGluLys 933
::: |||::: |||::: |||::: |||
3117 AGCAGAGTAAGCCCTTGGAAACAACATGGAAGAGTTCCACAGAAAGAAA 3166
934 GluSerLeuLeuGlnGluTyAsnAlaLeuLysGlnLysLeuAspGluG 950
::: |||::: |||::: |||::: |||
3167 GAAGATTTAAATCAACAAAAAAATACAGCAAAATAGCAGCAGAGTAA 3216
950 uTyHisGlnIleuAlaGluArgLysArgGluPheGlnGlnIleuLeuSPa 967
::: |||::: |||::: |||::: |||
3217 ACTGAATATATTTAAAGAGAGTTAAAAATATTCATGCTATATG... 3262
967 laLeuGlyArgLeuAsnMetLysIleLysGlyTyLeuAspSerLysLys 983
::: |||::: |||::: |||::: |||
3263 .....AAAGACATTTGAGATTTATATCAAGATGGGAAA 3295
984 AsnGlu...LysLeuLysGluLeuGlnGluArgHisValLeuCysHis 998
::: |||::: |||::: |||::: |||
3296 GACGACTTAAGAGCAAAAGAAAGAACTGATTAATAAGTAAATA... 3340
998 sSerGlnLeuGlnSerCysMetAlaLysGlnGlnArgIleSerAlaGluL 1015
::: |||::: |||::: |||::: |||
3341 ...GCTCAACTAGTGAATGCCAGAAACACAAAGAAAGATAATAGAGATA 3389
1015 euAsnLysSerLysGluLeuLeuGlnGlnGlnGlnGlnLysArgAsn 1031
::: |||::: |||::: |||::: |||
3390 TGACACTCATAGACAGAGATTTGATACACAGAGATACAAAGAAAGGGG 3439
1032 lLeaSPaSPLeuLysTyArgLysThrLysAlaAspValGluGlnLe 1048
::: |||::: |||::: |||::: |||
3440 CTACAGATTAACCTTACTTTAGAAAGAAAGAAATAGCAAGCTTAAAGAGT 3489
1048 uThrArgSPILeGluSerLeuGluArgLysLeuLeuSerIleGlySerL 1065
::: |||::: |||::: |||::: |||
3490 TGAA...GAGAAAGAAAGAAACAAATTTGAGAGAAATGGGTCAA 3530
1065 euSerAlaIleGluAlaAspLeuLysArgHisSerGlnGluLys... 1079
::: |||::: |||::: |||::: |||
3531 TGCAGGTTTCCAATGAAAGTGAACATCGAAGATGGAGAGAAACATA 3580
1080 GluArgLeuAsnSerGluPheAsnArgTrpGlnGlnIleuSerValTy 1096
::: |||::: |||::: |||::: |||
3581 GACATTTAAAGAAACAAATCATTAATTTGCCATTAGGCGACAGAAAGTTA 3630
1096 rGlnSerAsnIleSerLysHisLysGlnGluLeuLysLeuSerGlnTyL 1113
::: |||::: |||::: |||::: |||
3631 TGAAGAAAGAAATTTATCATTTTAAAGAAAGAACTCGAGAACCAATATTC 3680
1113 ysaSPILeGluLysArgTyThrAsnGlnPheLeuGlnLeuLysThrThr 1129
::: |||::: |||::: |||::: |||
3681 GGGATGCTGAGCAAGAAAGATAGAAATGATGATGTTATGAGACACAA 3730
1130 GluMetAlaAsnLysAspLeuAspArgTyThrThrAlaLeuAspLysAl 1146
::: |||::: |||::: |||::: |||
3731 GAACCTTGAAACAAGAGCTGATTTATTAAGAACTTGACCAAGG 3780
1146 alEuMetArgPheHisSerMetLysMetGluGluIleAsnLysIleIle 1163
::: |||::: |||::: |||::: |||
3781 AATATGAAATTTCCACAGTATGAAATGGAAGAAATCAATAAATTATAC 3830
1163 ysgLLeuLeuTrpGlnGlnThrTyArgGlyGlnAspIleAspTyIleSer 1179
::: |||::: |||::: |||::: |||
3831 GTGACCTGTGGCAAGTACCTATGCTGACCAAGATATGATATACATAAA 3880
1180 lLeaSPaSP...SerGluGlyAlaGluThrArgSe 1191
::: |||::: |||::: |||::: |||
3881 ATACGCTGTGATGCCGATGAAATGATACACCTTCTGATAAAGCGGAA 3930
1191 rTySerTyArgValAlaMetGlnThrGlyAspAlaGluLeuGluMetA 1208
::: |||::: |||::: |||::: |||
3931 TTATTAACCTACCGAGTGTGATGCTGAAGGAGACACACACCTTGATATGC 3980
1208 rGlyLysArgCysSerAlaGlyGlnLysValLeuAlaSerLeuIleLysArg 1224
::: |||::: |||::: |||::: |||
3981 GAGAGCATGACAGTGTGCAAAAGATTAAGCTTCATCATATTCGC 4030
1225 LeuAlaLeuAlaGluThrPheCysLeuAsnGlyIleLeuAlaLeuAs 1241
::: |||::: |||::: |||::: |||
4031 CTGGCCCTGGGTGAACCTTCTGCTCAACTGTGGCAGCATCATCTGCTGA 4080
1241 pGluProThrThrAsnLeuAspGlyProAsnAlaGluSerLeuAlaAla 1258
::: |||::: |||::: |||::: |||
4081 TGAGCAACAAACAAATCTTGACCGAGAAAGAAACATGAACTTTCACACTG 4130
1258 laLeuLeuArgIleMetGluAlaArgLysGlyGlnGlnLysAsnPheGlnLeu 1274
::: |||::: |||::: |||::: |||
4131 CTCTGTGTGAGATATAAAGCTGCTCAGCAGCGCTTAACCTTCAGCTT 4180
1275 lLevalIleThrHisAspGluArgPheAlaHisLeuIleGlyGlnArgG 1291
::: |||::: |||::: |||::: |||
4181 CTGGTAATCACTCATGATGAAGATTTGTGGAGCTTTTAAAGACGTTCTGA 4230
1291 nLeuAlaGluLysTyTyArgValSerLysAspGluAsnGlnHisSerI 1308
::: |||::: |||::: |||::: |||
4231 ATATGTGAGAAATTTCTACAGATTAAGAAACATCATCATAGTGTCTAG 4280

```

1308 lelle 1309
4281 AGATT 4285

seq_name: gb_pat1:A86963

seq_documentation_block: 5893 bp DNA PAT 22-JAN-2000
LOCUS A86963
DEFINITION Sequence 54 from Patent WO9838306.
ACCESSION A86963
VERSION A86963.1 GI:6735747
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 5893)
AUTHORS Dolganov,G.
TITLE TRANSCRIPTS ENCODING IMMUNOMODULATORY POLYPEPTIDES
JOURNAL Patent: WO 9838306-A 54 03-SEP-1998;
GENELABS TECH INC (US)

FEATURES
source 1..5893
/organism="unidentified"
/isolate="RAD50_SEQ"
/db_xref="taxon:32644"
389..4327
/note="unnamed protein product"
/codon_start=1
/protein_id="CA669369.1"
/db_xref="GI:6735748"
/translation="MSRIEKMSTLVGRSGIEDKDKQITFEFSPILITLVGNAGKTT
IIECLKTYICTGDEPPTGKNTFVHDKVAQETDVRAQIRLQFHDVNGELIAYGRSMVC
TQSKTEKTELEGVITRTKHGEKVLSSKCAIDEMISLSDVSKAVLNNVFCHQE
DSWMPISDEGAKOKDFEISATRYIKALETTLPOVOTOGQKEVOMEYLYKOYE
KACEIRDTQSKKACOLESSKEIYKSYENEDPLKKNLKEIENHLSKMKLDNEKALD
SKRQKMDKNSLEEKMEKYFOCTDEQLNDLNNHRTVNEKERKLYDCRRELEKLNK
ESRLNQEKSSELLVEGRQLQADRHQEHTRADSLQIDELDKTGTGRLIEKSELL
IKNFHLVBEROGKAKTANQNLNDEPAKETTLOKQIDELDKTGTGRLIEKSELL
SKRONELKNVYELQLEGGSDRIELDQELIAERLISAEKNSNVEYELKMEYISLQ
NKEKADLDRLRLKIDQEMEOINHTTTRTOMEMLTOKADDEQIRIKTSKSDSLTSL
LGTFPNKKOLEDLHSKSEINOTRDLAKLNKELASPEONKHNNELKRRBOSS
YEDKLTQVGSQDFESDLRLKEIEKSSKQRMALGATVYSQITQLTDENOSCP
VCCRVPTAELEQVLSLQSLRLAPDKLSTESLEKKEKRDMLGLVPMQSTII
DLKEKEIPELRLNKNVNRDILQRLKNDIEOEETLGLIMPEESAKVCLDVIIMERF
OMELKVEREIAQAKLOGIDLFTVQVNOKEOEKQELDTVSSKTELRKLIQOQ
OEOIOLKSTTNELKSEKLOISTNLOROOLEQTVELSYEOLVREIDABOVSP
LETTLEKPOEKELINKKRTSKIKADKINDIKKRYKNHGVKQIDENYIOGCKDXY
KKQKTELKNVIALQLESEKHKERINEMKRLMKQDIDTQRIQERWLDNLTLRKNEE
LKEVEBERQOHLKEMQOMVLOKMSHOKLEENIDNTRKNNALGRQKQVEEIIHF
KKELEBERQOHLKEMQOMVLOKMSHOKLEENIDNTRKNNALGRQKQVEEIIHF
IRDLMSYRGODIEYIEIRSDADENYASDKRRNRYRVMLKMGDPLMRGSCSAG
OKVLASLIIRLALAEFFRCNGLIIDEPTNLDRENIESLAHALVLIIRSRGQRNF
QLVLTTHDEDFVELGRSEVEKFRYIKKNIIDOCSELYKCSVSLGNVH"

CDS

BASE COUNT 2078 a 1098 c 1303 g 1414 t
ORIGIN

alignment_scores:

Quality: 1729.00 Length: 1385
Ratio: 1.919 Gaps: 38
Percent Similarity: 65.054 Percent Identity: 32.419

alignment_block:
US-09-538-396-2 x A86963 ..

Align seg 1/1 to: A86963 from: 1 to: 5893

1 MetSerThrValAspLysMetLeuIleLysGlyIleArgSerPhe..... 15
||||| :||||| :||| :||||| :||||| :
389 ATGTCCCGGATCGAAGATGACATCTCTGGCGCGGAGTTTGGAAAT 438
16 AspProAspAsnLysAsnValIleThrPhePheLysProLeuThrLeuI 32

439 AGAGCAAGAAAGATTAAGCAATATACATCTTCTTCAAGCCCTTACAAATTT 488
432 leValGlyProAsnGlyAlaGlyLysThrThrIleIleGlyLysLeuLys 48
489 TGGTTGGACCAATGGGCGGAAAGACACCATCATTTGAATGCTTAAAA 538
449 LeuSerCysThrGlyGluLeuProProAsnSerArgSerGlyHisThrPh 65
539 TATATTTTACTGAGGATTTCCCTCTGGAACCAAA...GGAAATACATT 585
65 eValHisAspProLysValAlaGlyGluThrGluThrLysGlyGlnIleL 82
586 TGTACACGATCCCAAGGTGCTGCAAGAAACAGATGTGAGACCCGATTC 635
82 yLeuArgPheLysThrAlaAlaGlyLysAspValValCysIleArgSer 98
636 GCTGCAATTCGTGATGTCAATGAGAACTTATGCTGTGCAAGATCT 685
99 PheGlnLeuThrGlnLysAlaSerLysMetGluPheLysAlaIleGluSe 115
666 ATGGTGTGTACTCAGAAAAGCAAAAGACAGAAATTTAAACTCTGGAAG 735
115 rValLeuGlnThrIleAsnProHisThrGlyGluLysValCysLeuSer 132
736 ACTCATTT...ACTAGAACAAAGCAT...GCTGAAAGGTGACGTGACCT 779
132 yArgCysAlaAspMetAspArgGluIleProAlaLeuMetGlyLysIse 148
780 CTAACTGTGCAAGAAATGACGAGAAATGATCAGTCTTGGGGGTTC 829
149 LysAlaValLeuGluLysAsnValIlePheValHisGlnAspGluSerAsn 165
830 AAGGCTGTGCTAAATTAATGCTATTTTCTGCTACAGAAAGATCTTAATG 879
165 ProLeuGlnAspProSerThrLeuLysLysAspPheAspIlePheS 182
880 GCTTAAATGAGGAAAGGCTTGAAGCAAAAGTTTATGAGATTTT 929
182 eValThrArgThrThrLysAlaLeuGluValIleLysLysLeuHisLys 198
930 CAGCAACAAAGATACATTAAGCTTGAAGACCTCGGAGGTAGCTCAG 979
199 AspGlnMetGlnGluIleLysThrPheArgGlyLysLeuGluAsnLeuG 215
980 ACACAAAGGCAAGAAAGTAAAGAAATATCAAAAGGACATAAATATCTAA 1029
215 nThrValLysAspGlnAlaHisLysLeuArgGluAsnIleAlaGlnAsp 232
1030 GCAATATTAAGGAAAGGCTGTGAGATTCGTGATCAGATTACAAATTAAG 1079
232 GlnGluLysSerAspAlaSerLysSerGlnMetGlnGluLysGlyLys 248
1080 AAGCCACAGTAAACATCTTCAAAAGAAATGTCAATTCATAGAAATGA 1129
249 IleCysGlyThrGluArgGluIleLeuGlnMetGluThrSerLeuAspG 265
1130 CTGTGATTCATTAAGAAATGCTTAAAGAAATGAATGAATCACTCTTA 1179
265 uLeuArgArgLeuGlnGlnIleAspIleLysAlaThrGluArgSer 282
1180 AATTAATGAACCTTGACATGAATTAAGCCCTTGATACCCAAAGAAC 1229
282 hrLeuThrGlnGlnHisGlyLysLeuAlaAlaLeuSerGlnGluAsn 298
1230 AATGAGAGAAAGATATAAGTGAAGTGAAGAAATGAAGAAAGTTT 1279
299 GluAspThrAspGlnGluLeuMetGluThrGlnThrLysPheGlnGluArg 315
1280 CAAGGACTGATGACCACTAATATGATATATCAATCAACACAGAAC 1329
315 gIleAlaLeuLeuGluThrLysIleSerLysLeuValArgAspMetAsp 332

1330 AGTAAGGAGAAAGAAAGAAATGTGACTGTATCGTGAACCTGGAA 1379
332 spglualasertysersevalleuserlysglnanserleuThr 348
1380 AACATAATTAAGAAATCTAGGCTTCTCAATCAGGAAATACGAACTGCTT 1429
349 Htsglulegilylsleuglnalaglnalalaspalahtleuthrmetly 365
1430 GTTGAAACAGGGTCTCTACAGCTCCAGACAGATGCCATCAAGAACAATAT 1479
365 shlsgluaraspseraspilelysaenlleCysThrlyshisnleug 382
1480 CCGAGCTAGAGATTCATTAATTCAGTCTTGGCAACACAGCTAGAAATGG 1529
382 lypProvalPProgluhisPropherthrasnspvalalameAspleuThr 398
1530 ATGGCTTGGACCTGGGCCATTCAGTGAAGACAGATTAATAATTTTCAC 1579
399 AsnArgilelysalArg.....LeuSerSerle 408
1580 AAATCTGTGAGAGAGACAGCAAGAGGGAAGCAAACTGCCAACCACT 1629
408 ugluAsnaspLeuaspLysLysLysSerAsnLysaspLysaspv 425
1630 GATGAATGACTTTCAGAAAAGAGACTGTGAACAAACAGATAGAT. 1678
425 alleuThrPlyshisTyrleuLyslleasnlaaArgTyrSerGluValasp 441
1679GAGATAAGAGATTAAGAAAACCTGCACTGGGA 1708
442 GlyGlnleuglnserlyslleGluSerMetserGlylleuArgArgar 458
1709 AGAATAATTTGATTAAATCAGAAATCTTAAGT.....AA 1743
458 gLysaspLysGluLysgluaraspalaalaglnalaleuSerLysp 475
1744 GAACAGAGATGAGCTGAATAATGAGATATGATACAGAGTGAAG 1793
475 heanleuSerArglleaspLysgluarGlnHisMetGlnleuVal 491
1794 GATCTTCAGACGATTCCTGAACTGACAGCAGAGCTC...ATPAAAGCT 1840
492 GluArgLysThrleuAlaleuGlyLysArgLysPyrLysSerleilese 508
1841 GAACGTGAG.....TTAGCAAGCTGACAGAAAACAGCAATGAGA 1881
508 rGlnLysArgThrGluValTyrSerleuGlnLys..... 520
1882 AACCTTAATAATGAGATTAATAGTCTCAAAATGAAAGAACAGACTTAG 1931
521lleLysValleuLeuArgLysaspilleleasnArgasn 534
1932 ACAGAGACCTGCTGAACCTGACACAGAGATGAGAGCTTAACCAACAT 1981
535 AlaaspLysArgValLysleuGly...leuLysAspAlaleuGlnse 550
1982 ACAACAAACACCTACCCAAATGAGAGATGCTGACCAAGCAAAAGCTGACA 2031
550 rSerLysaspLysLeuasnGlnlileValasnGlnHisLysaspLyslel 567
2032 AGAT...GAACAAATCAGAAAATTAATCTAGGCACAGATGAATTA 2078
567 yslLysValleuArgGlyArgasnProhegluLysaspMetLysLysGlu 583
2079 CCTCAGCTGTTG...GGATATTTTCCCAACAAACAGCTTGAAGAC... 2122
584 lleasnGlnAlaphetPProvalAspLysgluLysasnGlnLysargse 600
2123TGG.....CTACATAG 2133
600 rLysSerGlnlualaglnGlnleuLysPheThrGlnserLysValr 617
2134 TAATCAAAAGAAATTAATCAG..... 2155
617 hrAspAlaArgGlnleuThrLysLeuArgArgaspMet..... 630
2156ACCAGGACAGACTTGCCTCAATTTGAACAGAACTAGCTTCACT 2200
631 AspAlaLysArgArgPheLeuaspSerLysLeu...GlnserileuGln 646
2201 GAGCAGATTAATAATCAATTAATTAATGAACTAAAGAAAGGAAGAGA 2250
646 nllSerAlaAsnValaspMetPheProLysValleu.....GlnAsp 661
2251 GTTGTCCAGATACGAAACAGAGCTGTGTGATTTGTGTGAGCCAGAT 2300
661 lametAsnLysArgaspGlnGlnLysArgleuGlnAsnPheAlaasnGly 677
2301 TTGAAAGTGAATTTAGACAGGCTTAAGAGAAATTTGAATAATCATCAAA 2350
678 MetArgGlnMetLeuAlaPro.....PheGln 686
2351 CAGGAGCCATGCTGCTGGAGCCACAGAGATTTACTCCAGTCACTAC 2400
686 uhlLysleuAlaArgLysasnHis...ValCysProCysGlyArgAlap 702
2401 TCAGCTAACAGAGCAAAACAGATCATGTGCCCCGTTGTGAGAGACTT 2450
702 he...ThrProaspGlnGlnLysaspLysPheValLysLysGlnArgMetGln 717
2451 TTCAGACAGAGGCTGAGTTACAAAGAGTCAATCAGTGTGATTCAGCTTAA 2500
718 AsnSerSerThrAlaGlnArgSerLysAlaLeuAlaMetGlnSerSer 734
2501 CTGCGAGATTTGCTCCAGATAAATCACTCAACAGAACAGAGCTAAAAA 2550
734 nalaglnAlaleuPheGlnleuAspLysLysleuArgThrileTyrAsp 751
2551 AAAGGAAAGCGGCTGATGAATGCTGGAGCTGTGCCATGAGGCAAA 2600
751 lAtyValLysleuValGlnGluThrlePro...leuAlaGluLys... 765
2601 GCATTAATTTGATTGAAGAGAGAGAAATCCAGAAATTAAGAAACAACTG 2650
766 ...AsnLeuasnGlnHisLeuAlaAspLysGlnLysalaglnAlaph 781
2651 CAGAAATTCATAGACAGATACAGCGCTTAAGAAACAGATAGAAACA 2700
781 easpAspLeuLeuGly..... 786
2701 AGAAACACTCTTGGGTACAAATATGCTGAAGAAAGAGTCCAAAGTAT 2750
787ValleuAlaHisValGlnMetLaspArgAspAla 797
2751 GCCTGACAGATGTACAAATTAATGAGAGGTTCCAGATGAACTTAAGAT 2800
798 ValGlu.....AlaLeuGlnInProThrAspTh 807
2801 GTTCAAAAGAAAATTTGCACAAACAGCAGCTAGCTACAGAAATAGAC... 2848
807 rtleaspArgHisValHisGlnleuGlnleuValLysGlnValGln 824
2849 TTAGATCGAACTGTCCACACAGTCAACAG...GAGAAACAG 2888
824 sPLeuGlnLysAlaLeuaspSerSerLysArgGlyValLysSerleuGln 840
2889 AGAAACAGCAACAGTGAACCA.....GTTTCTAGTACAT 2926
841 GluileGlnleuGlnleuAsnPheleuGlnArgThrArgAspThrleuI 857
2927 GAATGAAATGTAAGCTT..... 2944
857 evalGlnValaspaspLeuArgaspGlnHisArgMetLeuasnGlnAsp 874
2945ATACAGACAGCAGAGAAACAGATTCACATCTAA 2978

874 etserleuaglnvalargtrpHisasnaIaarglulysVallys 890
 2979 AAATACAAAC.....ATGACCTAAATCTGAGAACTTCAG 3016
 891 AlaserleuaglnvalargtrpHisasnaIaarglulysVallys 907
 3017 ATATCCACTATTTGCAACGTCGACGAACTGGAGGAGCTGCA 3066
 907 uLeuAlaGlnGluysglulysVallysVallysVallys 924
 3067 ATATACCTGAGTACGCTCTTGTACAGAGATAAGCATCTAAAG 3116
 924 luserleuaspProleu.....Serlysglnlys 933
 3117 AGCAGATGAGCTTGGAAACACATTTGAAAGCTCCAGCAAGAA 3166
 934 GluserleuaglnGlnGluysglulysVallysVallys 950
 3167 GAGCATTAATCAACAAAAATTCAGCAACAAATAGCAGCATTA 3216
 950 uTyHisGlnLeuAlaGlnGluysglulysVallysVallys 967
 3217 ACTGATGATATTAAGAGAGATTAATAATTCATGCTATAG.... 3262
 967 lalauGlyArgLeuAsnMetLysIleLysGlyTyLeuAspSerLys 983
 3263AAGCATTTGAAATTAATTCAGATGCGAA 3295
 984 AsnGlu.....LysLeuLysGlnGlnGlnGlnGlnGlnGln 998
 3296 GACGCTTATAGACCAACAACTGACCTTAATAAGTATA.... 3340
 998 sserGlnLeuGlnSerCysMetAlaLysGlnGlnGlnGlnGln 1015
 3341 GCTCACTAAGTAAATGCGAGAACACAAAGAAATTAATGAAGA 3389
 1015 euAsnLysSerLysGlnGlnGlnGlnGlnGlnGlnGlnGln 1031
 3390 TGAGACTCTGAGACAGATATTTATACACAGATTAACAAAGCTG 3439
 1031 lleAspAsnLeuLysArgLysThrLysAlaAspValGlnGln 1048
 3440 CTACAGATTAACCTTACTTAAGAAAGAAATGAGGAATTAAGA 3489
 1048 uThrArgAspIleGlnLeuLysGlnGlnGlnGlnGlnGln 1065
 3490 TCAG.....GAGAAAGAAACACATTTGAGGAAATGCGCAAA 3530
 1065 euSerAlaIleGlnAlaAspLeuLysArgHisSerGlnGln 1079
 3531 TGACGCTTTGCAAAATGAAAGTGAACATGAGAGTGGAGAGCA 3580
 1080 GluArgLeuAsnSerGlnLysAsnArgTrpGlnGlnGlnGln 1096
 3581 GACAAATTAAGAAATCATTAATTTGCGATTAGGCGCAGAAAG 3630
 1096 rGlnSerAsnIleSerLysHisLysGlnGlnGlnGlnGln 1113
 3631 TGAGAGAGAAATTAATTCATTTAAGAAAGCTTCGAGAAACCA 3680
 1113 yAspIleGlnLysArgTrpHisGlnGlnGlnGlnGlnGln 1129
 3681 GGGATCTGAGGAAAGTATAGGAAATGATGTTGTATGAGAGACA 3730
 1130 GluMetAlaAsnLysAspLeuAspArgTrpThrAlaLeuAsp 1146
 3731 GAACCTGTGACAAAGATCTGATATTTATTAAGACTTTTACCA 3780
 1146 aluMetArgPheHisSerMetLysMetGlnGlnGlnGlnGln 1163
 3781 AATTAATCAAAATTTACAGATGATGAAATGAGAAATCAATTA 3830
 1163 ysgLLeuTrpGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1179

3880
 3881 GTGACCTGTGCGAAGTACCTATGTCGACAAAGTATTAATCAAG 3880
 1180 lLeasnsrAsp.....SerGlnGlnGlnGlnGlnGlnGln 1191
 3881 ATACGCTGTGATGCCGATGAAATGTCATCAGCTTCTATAAGCG 3930
 1191 rTySerTrpArgValAlaMetGlnTrpGlnGlnGlnGlnGln 1208
 3931 TTATAACACCGAGTGTGATGTCGAGGAGGACACAGCCTTGAT 3980
 1208 rGlyArgCysSerAlaGlnGlnGlnGlnGlnGlnGlnGlnGln 1224
 3981 GAGCAGATGACGCTGCGAAGAAAGTATTAAGCTCATCATTCG 4030
 1225 LeuAlaLeuAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1241
 4031 CTGGCTGTGCTGAAAGCTTCTGCTCACTGCTGATCTGCTGGA 4080
 1241 PGIupProThrTrpAsnLeuAspGlyProAsnAlaGlnSerLeu 1258
 4081 TGAGCCACAAACAAATCTTGACCGAAGAAACATGAATCTTGC 4130
 1258 lAlaLeuArgIleMetGlnAlaArgLysGlnGlnGlnGlnGln 1274
 4131 CTCTGCTGAGATTAATAAGTCCCTCACAGCAGCAGTTCAGCT 4180
 1275 lleValIleThrHisAspGlnGlnGlnGlnGlnGlnGlnGln 1291
 4181 CTGGTAACTACATGATGAGATTTGTGGAGCTTTAGAGCTTCGA 4230
 1291 nLeuAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1308
 4231 ATATGTGAGAAATCTACAGATTAATAAGAAACATGATGCT 4280
 1308 leIle 1309
 4281 AGATT 4285
 seq.name: gb_pat1:AR048375
 seq_documentation_block: 5893 bp DNA PAT 29-SEP-1999
 LOCUS AR048375 Sequence 54 from patent US 5821091.
 DEFINITION AR048375
 ACCESSION AR048375
 VERSION AR048375.1 GI:5970718
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 5893)
 AUTHORS Doljanov, G.
 TITLE Method of identifying activated T-cells
 JOURNAL Patent: US 5821091-A 54 13-OCT-1998;
 FEATURES
 source 1..5893
 BASE COUNT 2078 a 1098 c 1303 g 1414 t
 ORIGIN
 alignment_scores: Length: 1385
 Quality: 1729.00 Gaps: 38
 Ratio: 1.919 Percent Identity: 32.415
 Percent Similarity: 65.054
 alignment_block:
 US-09-538-396-2 x AR048375
 Align seg 1/1 to: AR048375 from: 1 to: 5893
 1 MetSerThrValAspLysMetLeuIleLysGlyIleArgSerPhe..... 15

389 ATGCCCGGATCGAAAGATGAGCATCTGCGGCGTGGGAGTTTGGAT 438
16 .AspProAspAsnLysAsnValIleThrPhePheLysProLeuThrLeu 32
439 AGAGGACAAAGATAGCAAAATATACATTCTTCCAGCCCTTCAAAATTT 488
32 LeuLysGlyProAsnGlyValIleGlyLysThrThrIleIleGlyLeuLys 48
489 TGGTTGGACCCCATGGGCGGAGAAAGACACCATCATTTGAATGCTTAAA 538
49 LeuSerCysThrGlyGlyLeuLeuProProAsnSerArgSerGlyHisThrPh 65
539 TATATTTGTAATGAGATTTCCCTCGTGAACCAAA..GGAATATACATT 585
65 eValHisAspProLysValIleGlyGlyThrGlyThrLysGlyLeuIle 82
586 TGTACACATCCCAAGGTTGCTCAAGAAACAGAGTGTAGACCCAGATTC 635
82 yLysLeuArgPheLysThrAlaIleGlyLysAspValValCysIleArgSer 98
636 GTCTGCATTTTCGTGATGTCATGAGAGAACTTATAGCTGTGCAAAAGATCT 685
99 PheGlnLeuThrGlnLysAlaSerLysMetGluPheLysAlaIleGlyLeu 115
686 ATGGTGTCTACTCAGAAAAGCAAAAAGCAGAAATTTAAACTCTGGAAG 735
115 rValLeuGlnThrIleAsnProHisThrGlyGlyLysValCysLeuSerT 132
736 AGTCATT...ACTAGAACAAAGCAT...GGTGAAGAGTCACTGAGCT 779
132 yTrpGlyAlaAspMetAspArgGluIleProAlaLeuMetGlyValSer 148
780 CTAGGTGTCAGAAATGACCCGAGAAAGATCACTTCTTGGGGTTCC 829
149 LysAlaValLeuGluAsnValIlePheValHisGlnAspGlnSerAsnTr 165
830 AAGGCTGCTCAATTAATGTCATTTCTGTCATCAGAGATTTCTAATG 879
165 pProLeuGlnAspProSerThrLeuLysLysLysPheAspIlePheS 182
880 GCCTTAAGTGAAGAAAGCCTTGAAGCAAAAGTTGATGAGATTTT 929
182 eAlaIleThrArgTyrThrLysAlaLeuGluValIleLysLysLeuHisLys 198
930 CAGCAAGATACATTAAAGCCTTAGAACCTCGGACGATCGCTAG 979
199 AspGlnMetGlnGluIleLysThrPheArgLeuLysLeuGluAsnLeuG 215
980 ACACAAAGGTGAAAGTAAAGATATCAATTCGAACATAAATATCTGAA 1029
215 nThrValLysAspGlnAlaHisLysLeuArgGluAsnIleAlaGlnAsp 232
1030 GCAATATATGAGAAAGAGCTGTGAGATCGTGCAGATTCAGAGTAAG 1079
232 InGluLysSerAspAlaSerLysSerGlnMetGluGlnLeuLysGlyLys 248
1080 AAGCCCACTTACATCTTCAAGGAAATGTCAATCTCATAGAGATGAA 1129
249 IleCysGlyThrGluArgGluIleLeuGlnMetGluThrSerLeuAspG 265
1130 CTGATTCATTAAGAAATCGTCTAAAGAAATTAACATATCTCTCTAA 1179
265 uLeuArgTrpLeuGlnGlyGlnIleAspIleLysAlaThrGluArgSerT 282
1180 AATATGAACTGACATGAATTAAGAACTTGAATGAGCCGAAAGAAC 1229
282 hTrpLeuThrGlnGlnHisGlyLysLeuAlaIleLeuSerGlnGluAsn 298
1230 AAATGAGAGAAAGATTAATGTAAGTGAAGGAAAGAAAGGAGTTT 1279
299 GluAspTrpAspGluGluLeuMetGluTrpGlnThrLysPheGlnGluArg 315
1280 CAAGGACTGATGAGCAACTAAATGACTTATATCAATTCACAGAGAAC 1329

315 gIleAlaLeuLeuGluThrLysIleSerLysLeuValArgAspMetAsp 332
1330 AGTAAAGGAGAAAGAAAGAAATGTTAGACTGTCATCTGTAAGTGGAAA 1379
332 spGluAlaSerTyrSerSerValLeuSerLysGlnAsnSerGluLeuThr 348
1380 AACATAATTAAGATCTAGGCTTCTCAATCAGAAATAATGAACTGTT 1429
349 HisGluIleGlyLysLeuGlnAlaGluAlaAspAlaHisLeuThrMetLys 365
1430 GTTGAAACAGGTCGCTTACAGCTCAAGCAGATGCCATCAAGAACATAT 1479
365 sHisGluArgAspSerAspIleLysAsnIleCysThrLysHisAsnLeuG 382
1480 CCGAGCTAGAGATTCATTAATTCAGTCTTGGCAACACAGCTAAATTTGG 1529
382 LyrProValProGluHisProPheThrAsnAspValAlaMetAsnLeuThr 398
1530 ATGGCTTTGAGCTGGGCCATTCAGTGAAGACAGATTAATAATTTTCA 1579
399 AsnArgIleLysAlaArg.....LeuSerSerLe 408
1580 AAACCTGTGAGAGAGAGACAGAAAGGAGGAAACAAACTGCCAACCACT 1629
408 uGluAsnAspLeuLeuAspLysLysLysSerAsnGluAspGlnLeuAspV 425
1630 GATGAATGACTTTCAGAAAGAGACTGTGAACAAACAAACAGATAGT. 1678
425 alleuTrpLysHisTyrLeuLysIleAsnAlaArgTyrSerGluValAsp 441
1679GAGATTAAGATTAAGAAACCTGGAGTGGGA 1708
442 GlyGlnIleGlnSerLysIleGluSerMetSerGlyIleLeuArgArgAr 458
1709 AGATTAATTAATGATTAATAATCAGAAATCTTAAGT.....AA 1743
458 gLysAspLysGlyLysGluArgAspAlaIleGluValGluLeuSerLysP 475
1744 GAACAGAGATAGCTGAAATAATGTAAGTATTAATTCAGCAGTTGGAG 1793
475 hAsnLeuSerArgIleAspGluArgGluArgHisMetGlnIleGluVal 491
1794 GATCTTCAGACAGATTTCTTAAGCTGACGAGAGCTC..ATTAAGCT 1840
492 GluArgLysThrLeuAlaLeuGlyLysArgAspTyrAspSerIleIleSe 508
1841 GAAGCTGAG.....TTAAGCAAGGCTGAGAAACACCATATGTGA 1881
508 rGlnLysArgThrGluValTyrSerLeuGlnGluLys..... 520
1882 AACCTTAAATGAAGATTAATTAAGTCTCAAAATGAAGAACACACTTGA 1931
521IleLysValLeuLeuArgGluLysAspIleIleAsnArgAsn 534
1932 ACAGAGACCTCGTAACTTGAACAGGAGATGAGCAGTTAAACATCAT 1991
535 AlaAspGluArgValLysLeuGly...LeuLysLysAspAlaLeuGluSe 550
1982 ACAACACACGCTACCCAAATGAGATGCTGACCAAGCAAGCAAGTGCACA 2031
550 rSerLysAspLysLeuAsnGluIleValAsnGlnHisLysAspLysIle 567
2032 AGAT...GAAACAAATCAGAAATAAATCTAGGACACAGTATTAATTA 2078
567 yLysValLeuArgGlyArgAsnProPheGluLysAspMetLysGlu 583
2079 CTTACCTGTTG...GGATTTTCCCAACAAACAAACGCTGAAGAC... 2122
584 IleAsnGlnAlaPheTrpProValAspLysGluTyrAsnGluLeuArgSe 600
2123TGG.....CTACATAG 2133

/codon_start=1
 /product="RAD50 homologue hRAD50"
 /protein_id="CAA99729.1"
 /db_xref="GI:2687853"
 /db_xref="SPTREMBL:O43254"
 /translation="MLIFSVRDMFAKMSILGVRSGFIEDKQIIITFFSPDITLIVGN
 GAKRTITIELKTYICTGDPFGTNGKGFVNDIPVYAOETDVAQRIQLOFRVNGELIIVN
 QRMSCNQKSKTEFKTLEGVITRTKHGEVYSLSKCAETIDREMISSLGYSKAVLVN
 IFCHQDSNMPISGKALKQKDFEITSATYIKALETLVROVROGQGVBEIQMELVY
 LKQYKACACIIRQIITSKEAQLITSKRYEALNEDPLNRKLEIHNLSKIMKLDN
 EIALDSRRKQMKRDNSELEKMKVQGDQEDNDLYNHNORVREKERKLVDCHEE
 LEILNKESRLNDQKSELIVQGRLOAQHDEHIAQPSLQSLATOLEDFEGRLE
 PSEBQIKNPHKILVREBQEGEATANDLMPAKETKQKQIDETDKQIGRITIE
 LKSEILSKQNELKNKYEEQLQEGSSDRIETDOLIKARELSKAKESNVETLKM
 EYISLQNERKADIRLTAKLDQEMQHLNHTTTRQMKLTKDKDKQEQIIRKISRS
 DETLSLGYFPNPKQLEMDLHSSKSELINQNRDRAKLNKLKELASEQKRNIIINELER
 EEDLSYEDKLFVPGCGSDPESDIDRIKEIEIKSSKQRMALAGTAVYSQITQLEDE
 NOSCCPQCRVQFQTEALQEAISDLQSLKLPADKISTESELKAKKRDQMGLAP
 MRSQIDILKEKELPELANKLQNVNDIQRILKNQIEBQETLGTIMPEESAKYCLTDV
 TIMRRQMKELDKVERKINQQAQKLOGIDLDRTYQVQVQNEQDEQHKIDITYSKTELNR
 KLIQDQEQIQHLAKSTNELSEKIQISITVQKROQLEEVTEVSTEQVILNIEIKQA
 KEQVSPLETTLEKFOQKEELINKKNTSNIAQDRKNDIEKVNKHGVKYNQIDENHQ
 DQGDNDVWQKRETELNKYKIAQLESECEREKEDINEDYMRADODIPQKQJQEWLNDNLL
 RKNNEELTEVEEBSKQHLKMGOMQVLYQMSKSEHQLEENIDIRKNNLALGKQGYE
 EEIILHFKELREQOFRAEKYEDMMVMTTELVNDIYKKTQDLQIMKFSMKM
 EELINKLITRELBQSTYRGQDTEYIEIRSDAENYSASPKRKNYVRYVMKIDGPAIDNR
 GRSAGQKVLASLIRLALAEFTCLNGCITIALDEPTTINDRENESIAHLAVETIKSR
 SQQRNQLVLTIDDEDEVLLGSRSEYEVETRIKKNIDQCEIYKCVSSLGFNVH"
 BASE COUNT 1630 a 661 c 901 g 931 t
 ORIGIN
 alignment_scores:
 Quality: 1711.00 Length: 1375
 Ratio: 1.899 Gaps: 35
 Percent Similarity: 65.527 Percent Identity: 31.855
 alignment block:
 US-09-538-396-2 x HSRAD50 ...
 Align seg 1/1 to: HSRAD50 from: 1 to: 4123

```

6 LysMetLeuIleLysGlyIleArgSerPhe.....AspProAspAsnLys 20
   ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
76 AAGATGAGCAATTCGTGGCGCTGGAGATTGGAAATGAGAGCAAAAGATA 125
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
20 SasnValIleThrPhePheLysProLeuThrLeuIleValGlyProAsnG 37
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
126 GCAATATATCACTTCTCTCAGCCCTTACATATTTGGTTGAGACCCAAT 175
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
37 LysAlaGlyLysThrThrIleIleGlyCysLeuLysLeuSerCysThrGly 53
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
176 GGGCGGGAAAGACGACCATCATGTAATGCTAAATATATATTTGTCGTGA 225
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
54 GluLeuProAsnSerArgSerGlyLysThrPheValHisAspProLys 70
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
226 GATTTCCTCCCTCGAGAACCAA...GGAATATATATTGTAACGATCCCAA 272
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
70 SValAlaGlyLysThrGluThrLysGlyGlnIleLysLeuArgPheLysT 87
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
273 GGTTCCTCAAGAAACGATGTGAGACCCAGCATGTGTCGCAATTTTCGG 322
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
87 hrAlaIleGlyLysAspValValCysIleArgSerPheGlnLeuArgin 103
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
323 ATGTAAATGGAGAACTTATAGCTGTGCAAAAGATCTATGCTGTCTACG 372
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
104 LysAlaSerLysMetGluPheLysAlaIleGluSerValLeuGlnThrI 120
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
373 AAAGCAAAAAGACAGAAATTTAAACCCCTCGAAGAGATCAT...ACTAG 419
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
120 eAsnProHisThrGlyLysValCysLeuSerLysArgLysCysAlaAspM 137
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
420 AACCAAGCAT...GGTGAAGAAGTAGTGTGAGCTTCAATGATGTGACGAA 466
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```



```
2174 TCGCACTGCTCCAGATAAACTCAAGTCAAGATTCAGAGTCAAGGCTTAAAGG 2223
      |||
735 AlaGluAlaLeuPheGlnGlnLeuAspLysLeuArgThrIleTyrAspAl 751
      |||
2224 AAGGAAGAAGCGCGTGAATGATGCTGGAGCTTGGCCCATGAGGCAAG 2273
      |||
751 ATYValLysLeuValGlnIleThrIlePro...LeuAlaGlnLys... 765
      |||
2274 CATTAATGATTTGAAGAGAGAAATACCAATTAAGAAACAACAACTGTC 2233
      |||
766 AsnLeuAsnGlnHisLeuAlaAspLysSerGlnLysAlaGlnAlaPhe 781
      |||
2324 AGAATGTCAATAGACATACAGCGCTTAAGAAAGCATAGAAAGCA 2373
      |||
782 AspAspLeuLeuGly... 786
      |||
2374 GAACACTCTTGGGTACAAATATGCTGAAGAAAGTCCAAAGTATG 2423
      |||
787 ...ValLeuAlaHisValGlnMetAspArgAspAla 798
      |||
2424 CCGTACAGATGTTACAAATTTAGAGAGGTTCCAGATGAACTTAAGATG 2473
      |||
798 alglu...AlaLeuLeuGlnProThrAspThr 807
      |||
2474 TTGAAAGAAAAATTCACAAACAGACCTAAGCTACAAAGAAATAGAC... 2520
      |||
808 IleAspArgHisValHisGlnIleGlnLeuValLysGlnValGlnAs 824
      |||
2521 TTACATGCAACTGTCCAACAAAGTCAACAG...GAGAAACAGA 2561
      |||
824 PLeuGlnTyrAlaLeuAspSerSerGlyArgGlyValLysSerLeuGln 841
      |||
2562 GAACACAGCACAAAGTTAGACACA...GTTTCTAGTAAGATTG 2599
      |||
841 IuileGlnLeuGlnLeuAsnPheLeuGlnArgThrArgAspThrLeuIle 857
      |||
2600 AATGCAATCTGAAGCTT... 2616
      |||
858 ValGluValAspAspLeuArgAspGlnHisArgMetLeuAsnGlnLysPhe 874
      |||
2617 ...ATACAGAGCACAGACAGAAACAGATTCACACTCTTAA 2651
      |||
874 tSerSerAlaGlnValArgThrPheHisAsnAlaArgGlnLysValLys 891
      |||
2652 AAGTACACACA...AATGAGCTAAATCTGAGAAACTTCAGA 2689
      |||
891 lAserSerIleLeuGlnArgPheGlnLysSerGlnLysGlnLeuValLeu 907
      |||
2690 TATCCACTAATTTGCACAGCTGCTGACAGACGAGAGACAGACTGTGAA 2739
      |||
908 LeuAlaGlnLysGlnLysGlnLeuIleValGlnLysLysLeuLeuGln 924
      |||
2740 TTATCCACTGAATGCTGCTTTTACAGAGAGATTAAGATGCTTAAGA 2789
      |||
924 uSerLeuAspProLeu...SerLysGlnLysG 934
      |||
2790 GCAGGTAGCCCTTTGGAACAAACATTCGAAAGTCCAGCAACAAAG 2839
      |||
934 lUsrLeuGlnGlnLysTyrAsnAlaLeuLysGlnLysLeuAspGlnLys 950
      |||
2840 AAGATTAATACAAAAAATATACAAACAAACAAATATGACACAGATTA 2889
      |||
951 TyrHisGlnLeuAlaGlnArgLysArgGlnPheGlnGlnLeuAspAl 967
      |||
2890 CTGAATGATTAATAAGAGAGCTTAA... 2916
      |||
967 alLeuGlyArgLeuAsnMetLysIleLysGlyTyrLeuAspSerLysLys 984
      |||
2917 ...AATATTCATGCTATATGAAAGACTTGA 2947
      |||
984 snGlu...LysLeuLysGlnLeuGln 991
      |||
```


934 AAGCCAGTAGCTCTCGCAGGAATGTGACGTCTAGCAGATGAG 983
 249 ILeysglYhrGluArgGluIleLeuGlnMetCylThrSerLeuAspG 265
 984 CTTGAGCCATTGAAGATCGCTGAAAGAGATTGACACAACTGTCTAA 1033
 265 uLeuArgArgLeuGlnGlnIleAspIleYsAlaThrGluArgSer 282
 1034 AATATGAACTTGTATATGAAATTAAGCTTGAGACCGCAAGAAAC 1083
 282 hLeuLeuThrGlnGlnIleGlnIleLeuAlaLeuSerGluLeuAsn 298
 1084 AAATGAGAAAGATATACATGAAATTAGACAGAAATGAAAGATTTT 1133
 299 GluSprhrAspGluIleuMetGluThrGlnThrYsPheGluGluArg 315
 1134 CAAAGGACTGTGAGCAGCTAAATGACTGTATCACAATCCAGCAAC 1183
 315 gIleAlaLeuLeuGlnIleThrYsIleSerYsLeuValArgAspMetAsp 332
 1184 AGTAAGGAGAAAGAAAGCGCCCTGTAGACTGCCAGCAGAACTGGAGA 1233
 332 spGluAlaSerYsSerValLeuSerYsGlnAsnSerGluLeuThr 348
 1234 AGCTGAATTAAGAGCAGCGCTCTCAACCAAGAAAGCAGACTGCTT 1283
 349 hIleGluIleGluYsLeuGlnIleGluAlaAspAlaHisLeuThrMet 365
 1284 GTGAGAGAGGCTGTCTACAGCTACAGCAGCAGATCCATCAAGACAT 1333
 365 sHISGluArgAspSerAspIleYsAsnIleCySThrYsHisAsnLeu 382
 1334 CCGAGCCAGACACTATTAATTCAGTCTTGCCACACACTTGAATGG 1383
 382 LyrProValProGlnHisProPheThrAsnAspValAlaMetAsnLeuThr 398
 1384 ATGGTTTGAAGCTGAGACATTCAGTGAAGACAGATTAAATTTTCAT 1433
 399 AsnArgIleYsAlaArg.....LeuSerSerLe 408
 1434 GAACCTTGTGAAGAGAGACAGAAAGAACTAAATGCCACCCAGCT 1483
 408 uGluAsnAspLeuLeuAspYsYsYsSerAsnGluAspGluLeuAspY 425
 1484 CTTGAGGAGACTGCACACAAAGAACTCTGAACAGAGCAGCTTAT. 1532
 425 alleuThrYsHisTYrLeuYsIleAsnAlaArgTYrSerGluValAsp 441
 1533GAGCTGAGGACAGAGAGAGAGGAGGAGCTGGG 1562
 442 GlnGlnIleGlnSerYsIleGluSerMetSerGlyIleLeuArgArg 458
 1563 AGGAGGATTGAGTGAAGACAGAGATCTCAG.....AA 1597
 458 gLyAspYsGluYsGluArgAspAlaIleGluValGluLeuSerYsP 475
 1598 GAAGCAGAGGAGCTGGCGCAGTGAAGAGATGAGCTCAGCAGCTGAGG 1647
 475 heAsnLeuSerArgIleAspGluArgGlnArgHisMetGlnIleGluVal 491
 1648 GCTCTCCGACAGATTCTGGAACCTGGAACAGAGCTC...ACGAAAGCT 1694
 492 GlnArgYsHisLeuAlaLeuGluYsGluArgAspTYrAspSerIleIle 508
 1695 GAACGCTAA.....TTAAGCAAGCCGAGAAATATAGCAGCATAGA 1735
 508 rGlnLYsArgThrGluValTYrSer.....LeuG 518
 1736 AACCTTAAGCAGAGATTAATGAGCTCCAAATGAGAAAGCTGACCTGG 1785
 518 LueGlnLYsIleYsValLeuLeuArgGluYsAspIleIleAsnArgAsn 534
 1786 ACAGAGACTTGCGAAAGCTGAGACAGAGATGAGAGTAAACCATCAT 1835

535 AlaAspGluArgValYsLeuGly...LeuLYsYsAspAlaLeuGln 550
 1836 ACAACAAACCCGACACAGATGAGATGCTCACCAAGACAAACTGGACA 1885
 550 rSerLYsAspYsLeuAsnGlnIleValAsnGlnHisLYsAspYsIle 567
 1886 AGAT...GAACAGATCAGAAATAAAGTCTAGGCACAGTATGACTAA 1932
 567 yLYsValLeuArgGlyArgAsnProPheGluYsAspMetLYsGlu 583
 1933 CTTCACTGTTG...GGCTACTTCTCT.....AACCAAAAAACAG 1967
 584 IleAsnGlnAlaPheThrProValAspYsGluYsArgAsnLeuArg 600
 1968 CTTGAAGACTGGCTTCAATTCTAAATCCAAAGAAATTATATGACAGCAGGA 2017
 600 rLYsSerGlnGluAlaGluGlnIleLeuYsPheThrGlnSerYsVal 617
 2018 CAGACTGCCAAACTGAACAAAGACTA.....G 2046
 617 hAspAlaArgGluGlnLeuThrYsLeuArgArgAspMetAspAlaYs 633
 2047 CTTCACTGAACAAATAAATCATATTAATATGACTAAAGAAAG 2096
 634 ArgArgPheLeuAspSerLYsLeuGlnSerIleLeuGlnIleSerAla 650
 2097 GAAGAACAGTTGTCCAGTTATGAAGATAGCTATTTATGTTGTGTAG 2146
 650 nValAspMetPheProLYsVal.....LeuGlnAspAlaMetAsnLYsA 665
 2147 CCAAGATTGTGAAGATATTGGGCACACTTAAGAAAGAAATTAAGAA 2196
 665 rGAspGluGlnLYsArgLeuGlnAsnPheAlaAsnGlyMetArgGluMet 681
 2197 CTTCAAGACGAGCGAGCTGTGTCGACAGAGCCACAGCAGATTACTCCAG 2246
 682 LeuAlaProPheGlnHisLeuAlaArgLYsAsnHis...ValCYsProCY 697
 2247 TTC.....ATTACGAGCTGACAGATGAAGAAACCGACTGCTGCCAGT 2290
 697 sCYsGluArgAlaPhe...ThrProAspGluGluAspGluPheValYsL 713
 2291 CTTGACAGAGTATTTCAGACAGAGCTGAGTTCACAGAACTCATACCG 2340
 713 ySGLnArgMetGlnAsnSerSerThrAlaGluArgSerYsAlaLeuAla 729
 2341 ACTTGACGTCAAGCTGAGGCTGCTCCAGATAAATCAAGTCAACAGAA 2390
 730 MetGluSerSerAsnAlaGluAlaLeuPheGlnIleuAspYsLeuArg 746
 2391 TCGGAAGTGAAGAAAGAGAGCGCGCGCTGATGAATGCTGGGCTTGT 2440
 746 gThrIleTYrAspAlaTYrValLYsLeuValGluGluThrIleProLeu 763
 2441 GCCCGTAGGCAAGCATATTGATTGAAAGAAAGAAATACCAAT 2490
 763 IagLYsAsnLeuAsnGlnHisLeuAlaAspGluSerGlnLYsAlaGln 779
 2491 TAAAGAACCACTACAGAGTGTCAATAGACATACAGCCCTTAAGAT 2540
 780 AlaPheAspAspLeuGluYsValLeuAlaHisValGlnMetAspArg 796
 2541 GACATAGAACAGCAGGAACACCTTGGTGAACATATGCTTGAGAGAGA 2590
 796 PaLaValGluAlaLeuLeuGlnProThrAspThrIleAspArgHisVal 813
 2591 AATGCTAAAGTGTGCTGACAGATGTTACATCATGAGAGGTTCCAGA 2640
 813 IsgLueGlnGlnIleuValYsGluValGluAspLeuGluTYrAlaLeu 829
 2641 TGGAGCTGAAGAGATGTGAAGAGAAATTCACACAGCAGCAGCTAAGCTG 2690

JOURNAL
Submitted (22-DEC-1999) Biochemistry and Molecular Biology,
Louisiana State University Health Sciences Center, 1901 Perdido
St., New Orleans, LA 70112, USA
E-mail: Journals@lsuhsu.edu

FEATURES

Source

CDS

/organism="Rattus norvegicus"
 /db_xref="taxon:10116"
 156. 4094
 /function="DNA double-strand break repair and
 recombination"
 /function-telomere length maintenance"
 /note="similar to Saccharomyces cerevisiae Rad50; null
 mutation is lethal in murine embryonic stem cells"
 /codon_start=-1
 /protein_id="

BASE COUNT	1549 a	840 c	1116 g	939 t
ORIGIN				

```
alignment_scores:
  Quality: 1666.00
  Ratio: 1.871
  Percent Similarity: 64.174
  Length: 1404
  Gaps: 37
  Percent Identity: 31.197
```

alignment_block:

Align seg 1/1 to: AF218576 from: 1 to: 4444

```

1  MeSerThrValAspIysMetLeuIleGlyGlyIleAspSerPhe... 15
156 ATGTCCCGGATGCAAAAGATGACACTCTGGCCGGCGCAACTTTTGGGAT 20
16  AspProAspAsnLysAsnValIleThrPhePheLysProLeuThrLeuI 32
206 AGAGGATTAAGATTAACCAATATCTCTTCTTCAGCCCCCTCCACAAATT 25
32  LeValAlGlyProAsnGlyValAsnGlySerThrPheIleIleGluCysLeuLys 48
256 TGGTGGACCCCAATGGGGGGGGGAAAGACGACCATCATGTAATGCTTAAG 30
49  LeuSerCysThrGlyLeuLeuProProAsnSerArgSerGlyHisThrPhe 65
306 TATATTTGTACTGAGATATTTCCTCCTGGAAACAAA...GGAAATACATT 35
65  eValHisAspProLysValAlaIleGlyLeuThrGluThrLysAlaGlyIleLeu 82
353 TGGTCAATGATCCCAAGGTGTGCTCAAGAAACAGATGGTGGTCCCAAAATTC 40

```

[illegible]

375 IecysThrLysHisAsnLeuGlyProValProGluHisProPheThrAsn 391
::: ||| :::: ||||| :::::
1276 TGGACGACCATCTTGAATTGGATGCTTTTGACGCTGGACCATTCAGTAA 1325
392 AsValAlaMetAsnLeuThrAsnArgIleLysAlaArg 404
||| :::: |||
1326 AGACAGATTAAAACTTTCATGAACTGTGAGAGAGAGACAGAGAGAGA 1375
405 LeuSerSerLeuGluAsnAspLeuAspLysLysLys 418
||| :::: |||
1376 AGCTAAACTGCCAGACGCTCTTGAGCGACCTTACAGACAAAGAGCGC 1425
418 eAsnGluAspGlnLeuAsp ValLeuThrPlyHisTyrLeuLysIle 424
::: ||| :::: |||
1426 TGAAGCAGACAGATGATGATGATGAGGAGACAAAGAGCGGCTGGCG 1475
425 ValLeuThrPlyHisTyrLeuLysIle 433
::: ||| :::: |||
1476 AGGATGATGCTGCTGAGACGCGAGATCTGACGAAAGAGACAGCTGACT 1525
433 eAsnAlaArgTyrSerGluValAlaAspGlyGlnIleGlnSerLysIleGlu 450
::: ||| :::: |||
1526 GAGC AACGTGAGGAATGAGCTGCAGC 1551
450 erMetSerGlyIleLeuArgArgArgLysAspLysGluLysGluArgAsp 466
::: ||| :::: |||
1552 AGCTGAGGCGCTCTCGACAGAGATCTGAGCTGACGACGAGCTGACA 1601
467 AlaIaGluValGluLeuSerLysPheAsnLeuSerArgIleAspGluArg 483
||| ||| ||| ||| ||| |||
1602 AAACGGAACTGAACTAGCAAG GC 1627
483 gGluArgHisMetGlnIleGluValGluArgLysThrLeuAlaLeuGlyG 500
||| :::: |||
1628 TGAGAAAAATAGCAGCATAGAAACCTTAA 1658
500 IuaArgAspTyrAspSerIleLeuSerGlnLysArgThrGluValTyrSer 516
::: ||| :::: |||
1659 GCAGAAATACTAAACCTCAAGAGAGAAAGCGGAC 1694
517 LeuGlnGlnLysIleLysValLeuLeuArgGluLysAspIleIleAsnArg 533
||| :::: |||
1695 CTGACAGAGAACTGCGGAACCTGATCAGAGATGAGCAGTTAAACCA 1744
533 gAsnAlaAspGluArgValLysLeuGly LeuLysLysAspAlaLeuG 549
||| :::: |||
1745 TCATACAAACACCGCACACAGATGAGATGCTTACAAAGACAAAGCTG 1794
549 IuSerSerLysAspLysLeuAsnGlnIleValAsnGluHisLysAspLys 565
||| :::: |||
1795 ACAAAAGAT GAACAGATCAGAAAAATTAAGTCCAGCACAGTATGAA 1841
566 IleLysLysValLeuArgGluArgAsnProPheGlnLysAspMetLysLys 582
::: ||| :::: |||
1842 CTAACTTCTCTGTTG GGATATTTTCT AACAAAA 1876
582 sGluIleAsnGlnAlaPheTyrProValAspLysGluTyrAsnGlnLeuVal 599
||| :::: |||
1877 ACAGCTTGAACAGCTGCTTCAATTCATTAATCCAAAGAGATTAATCAGACA 1926
599 rGserLysSerGlnGluAlaGlnGlnGluLeuLysPheThrGlnSerLys 615
||| :::: |||
1927 GCGACAGACTGCCAAACTGAACAAAGACTA 1958
616 ValThrAspAlaArgGluGlnLeuThrLysLeuArgArgAspMetAspAl 632
::: ||| :::: |||
1959 GCTTACGCCGACAAATTAATTAATTAATGACCTTAAGAA 2005
632 aLysArgArgPheLeuAspSerLysLeuGlnSerIleLeuGlnIleSerA 649
||| :::: |||
2006 AAAGAAAGACGACGCTGCTAGTTATGAGATTAATTAATGCTTTGATGTTGG 2055
649 IaAsnValAspMetPheProLysVal LeuGlnAspAlaMetAsn 663

2056 GTAGCAAGATTTGAAAGTAGCTAGACAGACTTAAAGAAATATGAA 2105
::: ||| :::: |||
664 LysArgAspGluGlnLysArgLeuGluAsnPheAlaAsnGlyMetArgG 680
||| :::: |||
2106 AAATCTCAAGACGCGAGCTGCTGCTGAGCAGACAGAGCTTACTC 2155
680 uMetLeuAlaProPheGlnHisLeuAlaArgLysAsnHis ValLysP 696
::: ||| :::: |||
2156 CCAAGTTT ATCACTCAGCTGACAGATGAAGAACAGCTCTGGCC 2199
696 rOCysCysGluArgAlaPhe ThrProAspGluGluAspGluPheVal 711
||| :::: |||
2200 CAGCTGTCAAGAGATTTTCAAGCGAAGCTGATTAACAGAAAGTATC 2249
712 LysLysGlnArgMetGlnAsnSerSerThrAlaGluArgSerLysAlaLe 728
::: ||| :::: |||
2250 AGTCACTTGCACTCCAGCTGAGGCTGCTCCAGATTAATCAAGTCAAC 2299
728 uAlaMetGluSerSerAsnAlaGluAlaLeuPheGlnGlnLeuAspLysL 745
||| :::: |||
2300 AGAATCAGAACTTAAAAAAAGAGCGCGCTGATGAATGCTGGCGC 2349
745 euArgThrIleTyrAspAlaTyrValLysLeuValGluGluThrIlePro 761
||| :::: |||
2350 TTGTGCCCATGAGGCAAGCATTAATTTGATTTGAAGAAAGAAATACCA 2399
762 LeuAlaGluLysAsnLeuAsn GlnHisLeuAl 772
||| :::: |||
2400 GAATTAAGAAACAGACTGACAGATGTCATATGACATACAGCCCTTAA 2449
772 AspGluSerGlnLysAlaGlnAlaPheAspLeuLeuLys 786
||| :::: |||
2450 GAATGACATTAAGAGAGACAGAGACA CTCTGGGTACAGTGA 2490
787 ValLeu 788
2491 TGCCGAAGAGAAAGTCTAAAGTCTCTGACAGAGCTACAGATCATG 2540
789 AlaHisValGlnMetAspArgAspAlaValGlu 799
||| :::: |||
2541 GAGAGCTTCCAGATGAGCTTAAAGACCTGAAAGAAATTTGCACACA 2590
800 AlaLeuLeuGlnProThrAspThrIleAspArgHisValHisGlu 815
||| :::: |||
2591 GGCAGCTTAAGCTGCGAGGGGTAGAC TTGATCGGCTCTCCAGCGG 2637
815 IeGlnGlnLeuValLysGluValGluAspLeuGluTyrAlaLeuAspSer 831
||| :::: |||
2638 TTAAACAG GAAAAACAAGAAAAACAACAACAACCTGATACA 2678
832 SerGluArgGluValLysSerLeuGlnGluIleGlnLeuGluLeuAsnPh 848
||| :::: |||
2679 GTTCCAGTAAGATTAATTAATTAATTAATTAATTAATTAAT 2711
848 eLeuGlnArgThrArgAspThrLeuIleValGluValAspLeuValG 865
::: ||| :::: |||
2712 ATACAG 2718
865 argGlnHisArgMetLeuAsnGluAspMetSerSerAlaGlnValArgTrp 881
||| :::: |||
2719 ACCAG CAGAACAAATTCAGACCTTAAAGAAAGTAAACA 2756
882 HisAsnAlaArgGluGluValLysAlaSerSerIleLeuGluArgPh 898
||| :::: |||
2757 AATGAGTGAATATCAGAACTGAGATGACCAATTTGCAAGCGCG 2806
898 eGlnLysSerGluGluGluValLeuLeuAlaGlnGluLysGlnGln 915
||| :::: |||
2807 TCAGCAAAATGAGAGACAGCTGGAATTTATCCACTGAATTCAGTCT 2856
915 euIleValGluLysLysLeuLeuGlnGluLysLeuAspPro 928
||| ||| |||

2857 TGACAGAGAAATAAGAGTCTAAAGACCAATTAACCTTTGGAGATA 2906
 929LeuSerLySGluGluSerLeuLeuGluGluTyrAs 941
 2907 GCACTGGAAAAGTTGCAGACAGAAAAGAAATGATCCACACAAAAA 2956
 941 nAlaLeuLySGluGluSerLeuGluGluTyrHisGluLeuAlaGluTyrL 958
 2957 TACACGTAACAAAATGGCTCAGGATTAAGATCAATGATATCAAAAGAAAG 3006
 958 yArGluPheGluGluGluGluLeuAlaLeuGluTyrGluLeuAsnMetLys 974
 3007 TTAATAATATCATGCTTATCATG.....AAAGAC 3035
 975 IleLySGluTyrLeuAspSerLyLysAsnGlu.....LysLeuLySGlu 989
 3036 ATAAGAAATTTATTTCAAGATGAAGAAAAGATGATCAATGAAGCAAAAAGA 3085
 989 uLeuGluGluTyrHisValLeuGluGluSerGluLeuGluSerGluMetL 1006
 3086 AACTGAACCTTAATAGATGTCATT.....CAACTTAATGATGATGACA 3129
 1006 lAlLySGluGluTyrLysSerAlaGluLeuAsnLySGluSerLySGluLeu 1032
 3130 AACCAAGAAAGAAATTAATAAGAAATGGAGAACCATGAGCAAGATATTT 3179
 1023 GlnGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 1039
 3180 GACACGAGAAAGATACAGAAAGGTGGTTACAGATTAACCTTACACTGAG 3229
 1039 gLySerThrLysAlaSerAlaGluGluGluGluGluGluGluGluGlu 1056
 3230 AAAAAGAGAGAGAGAACTTAAGAAAGTTGAA.....GAGCAACGAA 3270
 1056 lGluTyrGluLeuSerLysGluSerLeuSerAlaGluGluAlaSerLys 1072
 3271 AGCAAACTTTAAGAGATGGAGATGCAAGATTTACAGATGAAA... 3317
 1073 LysArgHisSerGluGluGluGluGluGluGluGluGluGluGluGlu 1084
 3318 AATGAGACCCAGAGTTGAAAGAGACATACACGATTAAGAGAAATCA 3367
 1085GluPheAsnArgTyrGluGluGluGluGluGluGluGluGlu 1099
 3368 CAGTTGGCATGGGGCGACAGAAAGC.....TATCAAGAG 3405
 1099 snIleSerLyHisLySGluGluLeuLySGluSerGluTyrLysAspLys 1115
 3406 AAATCTTCACTTAAGAGAGAGCTCGAGAACTCAGTTGCGGAGTGA 3455
 1116 GluLysArgTyrThrAsnGluPheLeuGluGluGluGluGluGluGlu 1132
 3456 GAGGAGAGATACAGAGAAATGATGATGATGATGATGATGATGATGAT 3505
 1132 aAsnLyAspSerLysAspArgTyrThrAlaLeuAspLysAlaLeuMetL 1149
 3506 GAATTAAGAGAGCTGACATCTTCTCAAAACCTTTCATCAGCAATTAATGA 3555
 1149 rGpPheHisSerMetLysMetGluGluGluGluGluGluGluGluGlu 1165
 3556 AATTTCACAGATGAAGAAATGAAGAAATCAACAAATTAATTCGATCTT 3605
 1166 TyrPheGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 1182
 3606 TGGCGGAGTACCTATCGTGGCAAGATATGAAATCAATGATGATGATG 3655
 1182 rAsp.....SerGluGluAlaGluGluGluGluGluGluGluGluGlu 1194
 3656 CGATGCTGAGAGAAATGATATCATGCTTCTGATTAAGCGCAATTAATCACT 3705
 1194 yTyrGluValMetGluGluGluGluGluGluGluGluGluGluGluGlu 1210
 3706 ACCGAGTGGTATGCTGAGAGGGGATACAGCCTTGAGCATGGCGGAGCA 3755

1211 CysSerAlaGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 1227
 3756 TGCAGTGTGACAGAGAGAGTCTGCTCTCTCATTCATCCAGCTGGCCCT 3805
 1227 uAlaGluThrPheCysLeuAsnGluGluGluGluGluGluGluGluGlu 1244
 3806 GCGGAAACCTTCTGCTGAGACGCTGGCATCTTCCCTTGATGAGCTTA 3855
 1244 hThrAsnLeuAspGluGluGluGluGluGluGluGluGluGluGluGlu 1260
 3856 CAACAAACCTGACCGAGAAACATGAGTCTGCGCACATGCTTGCTT 3905
 1261 ArgIleMetGluAlaArgLySGluGluGluGluGluGluGluGluGlu 1277
 3906 GAGATTTTAAAAAGCTGCTGCGACAGCGCACTTCAGCTTGCTGCTAT 3955
 1277 eThrHisAspGluArgPheAlaHisLeuIleGluGluGluGluGluGlu 1294
 3956 CACTCAGATGAAGATTTTGTGAGAGCTCTGAGACATCTGATATGTGG 4005
 1294 lLysTyrTyrArgValSerLysAspGluGluGluGluGluGluGluGlu 1310
 4006 AGAATTTCTACAGAGTGAAGAAACATCGACAGTGTCTGAGATGCTC 4055
 1311 SerGluGluGlu 1314
 4056 AAGAGCAGCATC 4067
 seq_name: gb_Pf4:AF057300
 seq_documentation block:
 LOCUS AF057300 4384 bp mRNA PRI 19-AUG-1999
 DEFINITION Homo sapiens truncated RAD50 protein (RAD50) mRNA, complete cds.
 ACCESSION AF057300
 VERSION AF057300.1 GI:5739042
 KEYWORDS
 SOURCE
 ORGANISM human.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 4384)
 Kim,K.K., Shin,B.A., Seo,K.H., Kim,P.N., Koh,J.T., Kim,J.H. and
 Park,B.R.
 TITLE Molecular cloning and characterization of splice variants of human
 RAD50 gene
 JOURNAL Gene 235 (1-2), 59-67 (1999)
 MEDLINE 99346140
 REFERENCE
 AUTHORS Kim,K.K., Shin,B.A., Kim,P.N., Seo,K.H., Koh,J.T. and Park,B.R.
 TITLE Direct Submission
 JOURNAL Submitted (03-APR-1998) Pharmacology, Chonnam University, College
 of Dentistry, 5 Hak-Dong Dong-Ku, Kwangju 501-190, South Korea
 FEATURES
 source 1. 4384
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="5"
 /map="5q31"
 /cell_line="Hela"
 1. 4384
 /gene="RAD50"
 /note="DNA recombination repair gene"
 674..4195
 /gene="RAD50"
 /note="RAD50-3: a 65-bp fragment compared to RAD50-2 is
 deleted by alternative splicing"
 /codon_start=1
 /product="truncated RAD50 protein"
 /protein_id="RAD50326.1"
 /db_xref="GI:5739043"
 /translation="MISSING:KAVLVNVIYFCQSDSNPLSEGRALKQKPEIRSAT
 RYKALETRQVRQYRGQGVKKEYQVMEKLYKKKACIRIQIISKEAQLSSKEIV

[illegible]

265 IuleuAArgIeuGInglYcIntleapIleySalatHrGuArSer 281
::: :|||:::||||
1047 AATTAAGAACTTGCAATGAATTAAAGCCTTGGATAGCGGAAGAAG 1096

```

282 ThrLeuLeuThrGlnClnHnSGlLysLeuAlaIleuSerGluClnAs 298
      ::          ::::  |||          :::::
1097 CAATGGAGGAAGATAAATACCTGCAACTGCAGAGCAAAAATGTGAAAAAGCTTT 1146

```

```

1147 TCAAGGACGTGATGAGCAACTAAATGACTTATATCAATCAACCAAGGAA 1196
      ::::: ||||| :::::
315 TGTlAlAlLeuLgUdThlYuslIseSrlYsLseuValAlAtAAsrMetAAsr 331
      ::::: ||| :::::
1197 CAGTAAAGGAGAGAAAGAAAGAAATGGTGAAGACTGCATCGTGAAGCTGGA 1246
      ::::: ||| :::::
332 AsrGulAlseSrlYsSerValLeuSrlYsGlnAsnSrglUeuth 348
      ::::: ||| :::::
1247 AATCTAAATTAAGAATCTAGGCTCTCATATCGAAGAAATAGAGACTCT 1296
      ::::: ||| :::::
348 rHlSglUleGlYsLseuGlnAlAlAlAlAsrAlAlAlYsLeuThrMetL 365
      ||| ||||| |||
1297 TGTTAAACAAGCTGCTCTCAACGCTCAAGCAAGTCCGATCAAGAACTA 1346
      ||| ||||| |||
365 YsHlSglUAtAAsrSerAsrIleYsAsnIleCysrThLYsHlAlAsnLeu 381
      ::::: ||| :::::
1347 TCCGAGCTAGAAATCATTAATTCAGTCTTGGCAACACACTGAATTTG 1396
      ::::: ||| :::::
382 GLYrProValrProGlnHlZrProHeHrHrAsrValAlAlMetAsrLeuth 398
      ::::: ||||| :::::
1397 GATGGCTTTGAAACCGTGAACATCTAGTGAAGACAGATTAATAATTTTCA 1446
      ::::: ||||| :::::
398 rAsnArTgIleYsAlArG.....LeuSerSertL 408
      ::::: |||
1447 CAATCTTGTGAAGAGAAACAAAGAGGGGAGCAAAAATCGCAACCAAC 1496
      ::::: |||
408 euGlnAsnAsrLeuLsArYsYsYsYsSerAsnGlnAsrPolLeuAsr 424
      ||| ||||| :::::
1497 TGATGAATGACCTTGCAGAAAGAGAGACTCGTGAACAAAGAACAGATAGAT 1546
      ::::: ||||| :::::
425 ValLeuTrPbYsHlYsLLeuYsAlAlAsnAlAlAtGTrSerGlnValAs 441
      ::::: ||||| :::::
1547 .....GAGTAAAGAGATGAAGAAATCGACTGG 1579
      :::::
441 polYglnIleGlnSrlYsIleGlnSerMetSerGlyLeuArGArGAs 458
      ||||| ||| :::::
1576 AAGAAATTAATGAGTTAAATTCAGAAATCTTAAGT.....A 1610
      :::::

```

```
1611 AGAAGCAGATGACCTGAAAAATCTGAGTATGATATACAGCAGCTTGGA 1660
      :::::::::::::: ::::: :::::
475 PheasnLeuSerArgIleaspGluArgGluArgHisMetGlnIleGluA 491
      :: ::::: ||| ::::: :::::
1661 GGATCTTCACAGAGATTTCTGAACCTGACAGAGAGCTC...ATAAAGC 1707
      :::::::::::::: ::::: :::::
491 LgluArgLysThrLeuAlaLeuGlyGluArgAspTyrAspSerIleLeu 508
      :::::::::::::: ::::: :::::
1708 TGAACGCTGAG...TTAAGCAAGCTGAGAAAAACAGCAATGTAG 1748
      :::::::::::::: :::::
508 ergLysArgThrGluValTyrSerLeuGlnLys... 520
      :: ::::: ||||| :::::
1749 AAACCTTAAAAATGGAAGTAAATAGTCTCCAAAATGAAAAAGCAGACTTA 1798
      :::::::::::::: :::::
521 .....IleLysValLeuLeuArgGluLysAspIleIleasnArgas 534
      ::::: ||| :::::
1799 GACAGACACCTGCGTAAACTTGACACAGAGATGGAGCAGTTAAACCATCA 1848
      :::::::::::::: :::::
534 nAlaaspGluArgValLysLeuGly...LeuLysLysAspAlaLeuGln 550
      ::::: ||| :::::
1849 TACAACACACAGTACCAAAATGAGATGTCACCAAGACAAAGCCTGACA 1898
      :::::::::::::: :::::
550 eSerLysAspLysLeuAsnGluIleValasnGlnHisLysAspLysIle 566
      ::::: ||| :::::
1899 AAGAT...GAACAATCAAGAAAAATTAATCTAGCGACAGTGAATTA 1945
      ::::: ||| :::::
1946 ACCTCACTGTG...GGATATTTCCACAAAAAACAGCTTGAAAGC.. 1990
      :::::::::::::: :::::
583 uIleasnGlnAlaPheTrpProValAspLysGluTyrAsnGlnLeuArg 600
      ::::: ||| :::::
1991 .....TGC.....CTACATA 2000
      :::::::::::::: :::::
600 eLysSerGlnGluAlaGlnGlnLeuLysPheThrGlnSerLysVal 616
      :::::::::::::: :::::
2001 GTAAATCAAAAGAAATTAATCAG..... 2023
      :::::::::::::: :::::
617 ThrAspAlaArgGlnGlnLeuThrLysLeuArgArgPheMet..... 630
      :::::::::::::: :::::
2024 .....ACCAGGACAGACTTGCCAATGTAACAAAGAACTGTCATC 2067
      :::::::::::::: :::::
631 .AspAlaLysArgArgPheLeuAspSerLysLeu...GlnSerIleLeuG 646
      ::::: ||| :::::
2068 TGACACAGAAATTAATCATATTAATGAACCTAAAAAGAAAGAGAGC 2117
      :::::::::::::: :::::
646 InIleSerAlaAsnValAspMetPheProLysValLeu.....GlnAsp 660
      :::::::::::::: :::::
2118 AGTTGTCCAGTTACGAAGACAGCTGTTGATGTTGTGTGACCCAGAT 2167
      :::::::::::::: :::::
661 AlameLysnLysArgAspGlnGlnLysArgLeuGlnAsnPheAlaAsnG 677
      ::::: ||| :::::
2168 TTGTAAGATGATTTAGACAGGCTTAAGAGAAATGAAAAATCATCAAA 2217
      :::::::::::::: :::::
677 yMetArgGlnMetLeuAlaPro.....PheG 686
      ::::: ||| :::::
2218 ACACGCGAGCATGTTGGCTGAGCCACAGCAGCTTACTCCAGTTCACTTA 2267
      :::::::::::::: :::::
686 LuHisLeuAlaArgLysAsnHis...ValSerProCysGlyAlaAla 701
      :::::::::::::: :::::
2268 CTCAGCTAACAGCAAAACAGCATGTTGCCCGTTGTGCAGAGAGTT 2317
      :::::::::::::: :::::
702 Phe...ThrProAspGlnGluAspGluPheValLysLysGlnArgMetG 717
      ::::: ||| :::::
2318 TTTCAGACAGAGGCTGAGTTACAAAGAGTCATGATTTGCGAGTCTAA 2367
      :::::::::::::: :::::
717 nAsnSerSerThrAlaGluArgSerLysAlaLeuAlaMetGlnSerSerA 734
      ::::: ||| :::::
2368 ACTCGAGCTTGCTCCAGATTAATCAACAGTCAACAGAAATCAGAGCTAAA 2417
      :::::::::::::: :::::
734 snAlaGluAlaLeuPheGlnGlnLeuAspLysLeuArgThrIleTyrAsp 750
      ::::: ||| :::::

2418 AAAAGAAAAAGCGCGTATGTAATGCTGGGACTTGTGCCATGAGCAA 2467
      :::::::::::::: :::::
751 AlaTyrValLysLeuValGlnGluThrIlePro...LeuAlaGluLys.. 765
      ::::: ||| :::::
2468 AGCATTAATGATTTGAAGAGAAAGAAATACCAAGATTAACAAACAACT 2517
      :::::::::::::: :::::
766 ....AsnLeuasnGlnHisLeuAlaAspGlnSerGlnLysAlaGlnAlaP 781
      ::::: ||| :::::
2518 GCAGATGTCAATGACACATACAGCCCTTAAGAACAGACATACAAAGAC 2567
      :::::::::::::: :::::
781 heAspAspLeuLeuGly..... 786
      ::::: ||| :::::
2568 AAGAAACACTCTTGTTGACAATAATGCCGTAAGAAAGAGTCCAAAGTA 2617
      :::::::::::::: :::::
787 .....ValLeuAlaHisValGlnMetAspArgAspAl 797
      ::::: ||| :::::
2618 TGCCGTGACAGATGTTACAAATTATGAGAGGTTCCAGATGCAACTTAAGA 2667
      :::::::::::::: :::::
797 aValGlu.....AlaLeuLeuGlnProThrAspT 807
      ::::: ||| :::::
2668 TGTTGAAAGAAAAATTCACACAAACAGCAGCTTAAGCTCAAGAGATAGAC 2716
      :::::::::::::: :::::
807 hIleAspArgHisValHisGlnIleGlnGlnLeuValLysGluValGlu 823
      ::::: ||| :::::
2717 ..TAGATCGAAGCTGCCAACAGTCAACGAG.....GAGAAACAA 2755
      :::::::::::::: :::::
824 AspLeuGlnTyrAlaLeuAspSerSerGlyArgGlyValLysSerLeuG 840
      ::::: ||| :::::
2756 GAGAAACAGCAGCAAGTTAGACACA.....GTTCTTACTAAGAT 2793
      :::::::::::::: :::::
840 uGlnIleGlnLeuGlnLeuAsnPheLeuGlnArgThrArgAspThrLeu 857
      ::::: ||| :::::
2794 TGAATGAATCTAGACTT..... 2812
      :::::::::::::: :::::
857 LeValGluValAspAspLeuArgAspGlnHisArgMetLeuAsnGluAsp 873
      :::::::::::::: :::::
2813 .....ATACAGACAGCAGAGAAAGATCAACATCTA 2845
      :::::::::::::: :::::
874 MetSerSerAlaGlnValArgTrpHisAsnAlaArgGlnGluLysValL 890
      ::::: ||| :::::
2846 AAAAGTACACA.....AATGAGCTAAAACTGGAACCTCA 2883
      :::::::::::::: :::::
890 sAlaSerSerIleLeuGlnArgPheGlnLysSerGlnGlnLeuValL 907
      ::::: ||| :::::
2884 GATATCCACTAATTTGCAACGCTGCTCAGCACTGAGAGAGAGCTGTG 2933
      :::::::::::::: :::::
907 euleuAlaGlnGlnLysGlnGlnLeuIleValGluLysLysLeuGln 923
      ::::: ||| :::::
2934 AATATTCACAGAGAGTCACTGTTGTACAGAGATTAAGATGCTAA 2983
      :::::::::::::: :::::
924 GluSerLeuAspProLeu.....SerLysGluLys 933
      ::::: ||| :::::
2984 GAGCAGGTAAGCCCTTGGAACAACATTGGAAAAAGTTCCAGCAAGAAA 3033
      :::::::::::::: :::::
933 sGlnSerLeuLeuGlnGlnLysAsnAlaLeuLysGlnLysLeuAspGln 950
      ::::: ||| :::::
3034 AGAAGAATTAATCAACAAAAAATACAGCACAACAAATACACAGAGTA 3083
      :::::::::::::: :::::
950 LuTyrHisGlnLeuAlaGluArgLysArgGlnPheGlnGlnLeuAsp 966
      ::::: ||| :::::
3084 AACTGATGATTAATTAAGAGAGGTTAAATATTCATGCGTATATG... 310
      :::::::::::::: :::::
967 AlaLeuGlyArgLeuAsnMetLysIleLysGlyTyrLeuAspSerLysL 983
      ::::: ||| :::::
3131 .....AAAGACATTGACAAATATTAATTCAGATGGGAA 3162
      :::::::::::::: :::::
983 sAsnGlu.....LysLeuLysGlnLeuGlnGlnLysArgHisValLys 998
      ::::: ||| :::::
3163 AGACGACTATTAAGAAACAAAGAACTGAACCTTAATTAATTAATTA 3208
      :::::::::::::: :::::
998 sSerGlnLeuGlnSerCysMetAlaLysGlnGlnArgLysIleSerAlaGlu 1014
      ::::: ||| :::::
3209 ..GCTCAACTAGTGAATGCCGAGAAACAAAGAAAGATTAATGAAGAT 3256
      :::::::::::::: :::::
```

1015 LeuasnlySerlySgluLeuLncInglngIngluInleuLysArgas 1031
 3257 ATGACACTCATGACAGAGATATTGATACACAGAGATACAGAAAGGCG 3306
 1031 nileaspasnlleuLysTyArgLysThrlYsAlaSPValgluInL 1048
 3307 GCTACAAATACCTTACTTTAGAAAAAGAAATAGACACTAAAAGAG 3356
 1048 eutHrArgAsPIlegluSerleugluArgleuLeuSerIlegluSer 1064
 3357 TTGAA.....GAGAAAGAAACAACATTGACAGAAATGGGTCA 3397
 1065 LeuSerAlaIlegluAlaSPleuLysArgHisSerGlnluLys.... 1079
 3398 ATCGAGTTTGCAAAATGAAAGTGAACATCAGAAAGTTGGAAGAACAT 3447
 1080 GluArgleuasnSerGluPheAsnArgTyrPheGlnluLysSerVal 1096
 3448 AGACATATATAAGAAATCATATTGGCATTAGGGCCAGAAAGGTT 3497
 1096 YrgInSerAsnIleSerLysHisLysGlnluLeuLysLeuSerGlnTyr 1112
 3498 ATGAGAGAAATATTTCATTTTAAAGAAAGACTCGAAGACCAATTT 3547
 1113 LysAspIlegluLysArgTyrThrAsnGlnPheleuGlnleuLysThr 1129
 3548 CGGATGCTGAGAAAGATATAGAAATGATGATGTTATAGAGAACAC 3597
 1129 RgluMetAlaSNLysAspLeuAspArgTyrTyrThrAlaLeuAspLys 1146
 3598 AGAATCTTGAAACAAGATCTCGATATTATATAGACTCTTGACCAAG 3647
 1146 IalMetLarPheHisSerMetLysMetGluGluIleasnLysIlele 1162
 3648 CAATATGAAATTCACACTATGAAATGGAAGAAATCATATAATTA 3697
 1163 LysGluLeuTyrPglInThrTyrArgLysGlnAspIleAspTyrIle 1179
 3698 CGTGCAGCTGGCGAAGTACCTAGTGCACAAATATGATACATAGA 3747
 1179 rIleasnSerAsp.....SerGlnluAlaGlnTyrArg 1191
 3748 AATAGGCTGTCGCGATGAAATGATATGACCTCTGATTAAGCGGA 3797
 1191 eTyrSerTyrArgValMetGlnThrGlyAspAlaGluLeuLumet 1207
 3798 ATTATTAACCGAGTGTGATGCTGAAGGAGACACACCTTGGATAG 3847
 1208 ArgGlyArgCysSerAlaGlyLysValleuAlaSerleuIleleAr 1224
 3848 CGAGACAGATGAGTGTGACAAAGATATAGCCTCATCATCATTCG 3897
 1224 GLeuAlaLeuAlaGluThrPheCysLeuAsnGlyIleleuAlaLeu 1241
 3898 CCTGGCCTGTGGAAGCTTCTGCTCAACGTGGCATTGCTGGTGG 3947
 1241 sPgluPThrThrAsnLeuAspGlyProAsnAlaGluSerleuAla 1257
 3948 ATGAGCCAAACAATCTTGACCGAAGAAACATTGAATCTTGGACAT 3997
 1258 AlaLeuLeuArgIleMetGluAlaArgLysGlnluAsnPheGlnle 1274
 3998 GCTCGGTGAGATATAAAGTCGCTCAGACGAGCGTAACCTCCAGCT 4047
 1274 uIleValIleThrHisAspGluArgPheAlaHisleuIleGlnluArg 1291
 4048 TCTGTATACACTCATGATGAAGATTTGTGAGCTTTAGACACGTTCT 4097
 1291 InleuAlaGluLysTyrTyrArgValSerLysAspGluAsnGlnHis 1307
 4098 AATATGTGAGAAATTTACAGGATTAAGAAACATCGATCAGTCTCA 4147

1308 Ilele 1309
 4148 GAGATT 4153
 seq_name: 'gb_PL3:SCRAD50'

seq_documentation_block: 4775 bp DNA PLN 12-SEP-1993.
 LOCUS SCRAD50
 DEFINITION Yeast RAD50 gene for 153 kD protein.
 ACCESSION X14814
 VERSION X14814.1 GI:4272
 KEYWORDS DNA repair; DNA-binding protein; meiotic recombination; RAD50 gene.
 SOURCE Saccharomyces cerevisiae.
 ORGANISM Saccharomyces cerevisiae.
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomyces.

REFERENCE
 AUTHORS Alani,E.
 TITLE Direct Submision
 JOURNAL Submitted (21-MAR-1989) Alani E., Harvard University, 7 Divinity
 Avenue, Cambridge MA 02138, U S A
 REFERENCE
 AUTHORS Alani,E., Subbiah,S. and Kleckner,N.
 TITLE The yeast RAD50 gene encodes a predicted 153kd protein containing a
 purine nucleotide binding domain and two large heptad repeat
 regions
 JOURNAL Genetics 112, 47-57 (1989)
 REFERENCE
 AUTHORS Raymond,M.E. and Kleckner,N.
 TITLE Expression of the Saccharomyces cerevisiae RAD50 gene during
 meiosis: steady-state transcript levels rise and fall while
 steady-state protein levels remain constant
 JOURNAL Mol. Gen. Genet. 238 (3), 390-400 (1993)
 DATA kindly reviewed (25-SEP-1989) by Alani E.

JOURNAL
 MEDLINE
 COMMENT
 FEATURES
 SOURCE
 Location/Qualifiers

misc_feature

CDS

1..4775
 /organism="Saccharomyces cerevisiae"
 /strain="REB21"
 /db_xref="taxon:4932"
 /chromosome="chromosome 14"
 /clone="pSG205"
 /note="region of transcription start"
 /note="153 kD protein (AA 1 - 1312)."
 /codon_start=1
 /protein_id="CAA32919.1"
 /db_xref="GI:4273"
 /db_xref="SWISS-PROT:P12753"
 /translation="MSAIVKLSIGIRSFDSNDRETFEGRPPLLIVGNGSGRTTII
 ECLKATATGDLPPRSKGVFIHPDKTIGEDIRAOVLAFSTANGNMYRNIDLM
 KRTTTPKTELEGOLVANNNGSDRSITSTRELDAOVLPLVGPALLLEVITVCHOD
 SLPFLPSNULKKFDFIRPAKFTKLDLTKSTKDKMSDILKLSQVPHLTKDNR
 SKAKMLIHOLQKIDYNEVESEIQLNETEKSDKESQDOKLISVENVLKN
 TKLISDQVRLNSIDILSKRPDLQNLNLSFSLMDNNQDLBLEDSILKRO
 SLOSLSNLSIRGSELAGETEKERNHLSSLEKAFQFGLNKEINSDSAQVNH
 EMSQKAFIISODLTDIDFQAKIDLETNLSDIKISIVSONLENNKROKLIH
 SEBAEKLFKFSLSNDSANHELENKTKREKLSMESNITIPKLNKKEENNMII
 ILENQIEKFDRIKTNQADLFAKGLIKSTYTKIDELQKTEKLNDSRPYRP
 LTFQFQRADEMFQKLFIMQKNIALNNKMEHLDRTYNALYNNTLEKDLQDNK
 SKREVILSENPEDCTIDEDYVEETELSYKTLLENLMQHTLEFRKLLEIAE
 RDSQCYLSKRFENESKSLQELKRTDANETKLDYQNEKEVLSRLLEKHI
 TLNSINKEIDNSQKLEKKEETKSKSLDELEVSTKDKSELAESIEIPLEK
 PLYIEKLKLENSKRTISELSYNSSEGIQVDELPOQRKNDSLBELKRTISD
 LQMKDEKVENSMNLINKEKELTVSEISSLGOKONIDSLRSKRNNIDDSVK
 ELKARIISLNKNDQASVIDKYNEDIDVRNKKQVADINLIDRFQTTIEVDFE
 EAKGFDELQTTIKELIELEKQMLKLEQDLKSNVEEKRLADSNNEKNKQMLE
 LIEKFSOLOHIESEISLDVONAEADKQESLSLRTPEKLSSENAKLEMQOL
 ONODISLHOLDRTDYDKERNYKHEWEIQTFRFVDIDVYSKALDSAIMKYHGLK
 ODRIITDELKRTYSGTDITIKIRSDPVSRYKGSYRYRYVMKQVDELMDRKO
 SAGKVLASIIIRLALSETFGANGVIALDEPTTNDEENIESLANSHTINMRKO


```

misc_feature      642..707      KNFQIVITHEDEKELGHNNAAFIDHEFKVRDRDROKSQIEWVDINRTVY"
                    /note="nucleotide-binding domain"
misc_feature      1089..1820
                    /note="heptad repeat"
misc_feature      2784..3542
                    /note="heptad repeat"
BASE COUNT      1827 a      768 c      895 g      1285 t
ORIGIN
alignment_scores:
    Quality: 1477.50      Length: 1404
    Ratio: 1.666      Gaps: 41
    Percent similarity: 63.177      Percent identity: 29.701
alignment_block:
US-09-538-396-2 x SCRAD50
Align seg 1/1 to: SCRAD50 from: 1 to: 4775
1 MetSerThrValAspLysMetLeuIleLysGlyIleArgSerPheAsp 17
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
558 ATGAGCGCTATCTATAATATCTATCAGGCGATACGGCTTTTGACTC 607
17 cAspAsnLysAsnValIleThrPhePheLysProLeuThrLeuValG 34
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
608 CAATGATAGGGAACCTATGATTTGGCAAGCCTGACCTTAAATGTCG 657
34 LProAsnGlyAlaGlyLysThrThrIleIleGlyLysLeuLysLeuSer 50
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
658 GCATGATAGTGTGAGTAAACAGCATATCATGATGTTAAAGTACGCT 707
51 CysThrGlyLysLeuProAsnSerArgSerGlyHisThrPheValH 67
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
708 ACCACAGGTATCTACCGCCCAACAGCAAGAGGAGA...GTATTCATTC 754
67 sAspProLysValAlaGlyLysThrGlyThrLysGlyIleLeuVal 84
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
755 TGACCCGCAATACCTGCTGAGAAAGCATTAAGCCTCAGCTCAACATCG 804
84 rPheLysThrAlaAlaGlyLysAspValAlaCysIleArgSerPheGln 100
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
805 CGTTTGAGATGCGCAATGACATATATGATGTCAACCAAAATATATCAG 854
101 LeuThrGlnLysAlaSerLysMetGluPheLysAlaIleGluSerVal 117
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
855 TTCTATATGAAAGAGACTACTACTACTATTAGACTTTAGAGCCAGT 904
117 uGlnThrIleAsnProHisThrGlyGlyLysValCysLeuSerThrArg 134
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
905 GGTGCGCTATTAAT...AATAGTGTGACCGCAGTACTTGTCCACCCGGT 951
134 yAlaAspMetAspArgGluIleProAlaLeuMetGlyValSerLysAla 150
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
952 CTCGTGAATTAAGCGCACAGTACCCCTGATCTGGAGAGTCCAAAAGCG 1001
151 ValLeuGluAsnValIlePheValHisGlnAspGluSerAsnTrpPro 167
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1002 ATCTTGGAATACGTGATATCTGTCAACAAAGAGACAGCTGTGGCCACT 1051
167 uGlnAspProSerThrLeuLysLysPheAspPheIlePheSerAla 184
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1052 GAGTGAACCCATCCAAATCTTAAAGAAAGTTGATGAATTTTCCAGCGA 1101
184 hArgGlyThrLysAlaLeuGluValIleLysLysLeuHisLysAspGln 200
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1102 TGAATTTAAAGGCTGACATTAATAATCCATTAAGAAAGACATG 1151
201 MetGlnGluIleLysThrPheArgLeuLysLeuGluAsnLeuGlnThr 217
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1152 TCCGTGATATCAAGTTAATAAGCAATAGTAGAACACTTAAAGTTG.. 1199

```

```

217 LysAspGlnAlaHisLysLeuArgGluAsnIleAlaGlnAspGlnLul 234
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1200 .....GATTAAGACA 1209
234 ySerAspAlaSerLysSerGlnMetGlnLeuLysGluLysIleCys 250
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1210 GANCAAGCAATGAAAGCTCAATATCCACCAATGCAAACTAAGATGAT 1259
251 GlyThrGluArgGluIleLeuGlnMetGluThrSerLeuAspGluLeu 267
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1260 CAGTATATGAAAGAGAGTGTGAATTTGAAATCTCAATGATGAATC.. 1307
267 ArgLeuGlnGlyGlnIleAspIleLysAlaThrGluArgSerThrLeu 284
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1308 .....ACTGAGAAATCCGATTAAC 1326
284 eu.....ThrGlnHisGlyLysLeuAlaAlaLeuSerGluGln 297
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1327 TGTTAATTAATCAAGATTTCCAAAATACTATCCAAAGTAGAGAAC 1376
298 AsnGluAspThrAspGluGluLeuMetGluTrpGlnThrLysPheGln 314
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1377 TTGAAAATCAAAACTCTCAATTAGTATCAAGTGAAGAGATTGTCAA 1426
314 uArgIleAlaLeuLeuGlu.....T 321
1427 CTCATATGATATTTCTCGATTTATCGAAGCCAGATTTCAAACTTATG 1476
321 hIleLysIleSerLysLeuValArgAsp.....MetAspAsp 332
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1477 CAAACTCTCTAAAGTCTGATGATGATGAATATACAGTTGAGAGATTG 1526
333 GluAlaSerTyrosSerValLeuSerLysGlnAsnSer.....GluLe 347
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1527 GAAACCGATTTTCAAGTTTGAAAGACCGTCACATCACTTGACAGACCT 1576
347 uThrHisGluIleGlyLysLeuGlnAlaGluAlaAspAlaHisLeuThr 364
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1577 TTCTAATTCCTCATCAGCGCGCAGCGCTGACTAGAGCT..... 1616
364 eLysHisGluArgAspSerAspIleLysAsnIleCysThrLysHisAsn 380
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1617 .....GTAAGAGACATACGAAAGAAAGCAAGAAATCACTTTCATCG 1658
381 LeuGlyProValProGluHisProPheThrAsnAspValAlaMetAsn 397
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1659 CTAAGAAACCCCTTCCAAACATAGTTC.....CAAGGCT 1693
397 uThrAsnArgIleLysAlaArgLeuSerSerLeuGluAsnAspLeuVal 414
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1694 TTCAAATATTGAAGACAGTATATGCTCAAGTAAATCAGCAATGTGCAC 1743
414 sP...LysLysLysSerAsnGluAspGlnLeuAspValLeuTrpLysHis 429
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1744 AGTCAAGCTTCATCCTCAGAGATTACTGACACTATAGATCAATTT 1793
430 TyrLeuLysIleAsnAlaArgTyrosGluValAspGlyIleGln 446
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1794 GCAAGAGATTTCACTCACTAAAGAAACAAATTTATCCCACTTGATTAAT 1843
446 rLysIleGluSerMetSerGlyLeuLeuArgArgLysAspLysGlu 463
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1844 AATAACCGTGTGATTCACAGATTTGGCAATTAATAAAGAGAGATCTA 1893
463 ys.....GluArgAspAlaAlaGluValGluLeuSerLysPhe 475
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1894 AATTAATTCATGACTCGAAGAGCTTGCTGAGAAATGAAATCATTTAA 1943
476 AsnLeuSerArgIleAspGluArgGluValHisMetGln...IleGluVal 491
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1944 AGTTGTCTCTCAAGATACCTCAATCATCAAGAACTGAAAACTTAAAC 1993
491 lGluArgLysThrLeuAlaLeuGlyGluArgAsp.....T 503

```


1994 ATACAAAGAAAGCTTCATCTTGGAAAGTGAAATATTATACCAAAAT 2043
503 yraspserlleieserleugllyasrghgluvaltyrserleuglly 519
2044 TGACACGAAATAGAGAGAAATATAGATGATATCTTAGAGAAAT 2093
520 lysllelysalaleuaurghglulysaspillellesnargasnlaas 536
2094 CAATTTGAGAAATTCAGATGATATGAAACGAAATCAACACCGGA 2143
536 pgluaurgvalysleugllyleuylsaspalaleugllyserleu 553
2144 CTTATACGCAACTAGACCTATTAAGAAGTCCATTAACTAAATTAG 2193
553 splysleuasnglulilevalasn.....glulhlyaspllylely 567
2194 AGAGATTCAAAAGATTACGAAATATACAGATGATCTAGAAATAG 2243
568 lysvalleuaurghgllyasrghglulys..aspmetlystsgl 583
2244 CAAGTATTTCCCTCACCAGGAATTTCAAGAGCTGATTTAGAAATGGA 2293
583 ulleasnglulalepethrprovalaspllysgllyrasngluleuaurgs 600
2294 TTTTCGAAATATTATCATATATGCAAAAGATATAGCATCACACAA 2343
600 erlyserleugllyalaglugllyleuylserghserleuylval 616
2344 AAAAGATGACGAATTCGATAG.....AGATAT 2372
617 thrasplalarglglulyleuthlyleuaurghargaspmetaspalaly 633
2373 ACAAAAGCTTTGTACAAATTTGAACACCATTTGAAAGATTTACAGATTA 2422
633 sargargpethleuaspserlyleugllyserleu.....leuglly 647
2423 TCGAAGATGAAAGAAAGATGATACACTACACGAAACTTACCAG 2472
647 leserlaasnvalaspmetpethrprolysalaleugllyasplametasn 663
2473 AGGATGCTACTATGATGATATATATGATGCTCTGAGAA.....ACG 2516
664 lysargaspglglulysargleugllyasnphelalasnlymetarggl 680
2517 GAGCTCTCTTATAGACTGCTGGAACCTTGAAA.....ATGCATCA 2560
680 umetleuallaprophglu.....hileuallarglyasn 693
2561 AACTACTTAGAATTTAATAGAAAGCATTAGATGCTGGAACGTGACA 2610
693 lsalvalcyspocyscysgluargalaphethrproaspglulysarglu 709
2611 GGTGCTGATTTGTCTAGAAATTT.....GAAATGATCG 2651
710 phevallysalys.....gluargmetgluasnserleuylalagluar 724
2652 TTTTAAAGTAAAGTTGTTACAGAACTAAACAAACAGACCCCAATTT 2701
724 gserlysalaleuallametgluaserasnlaagluallaleuphglng 741
2702 TGAATAAACTTTGAAAGACACTGTCGAAACGAAAG.....GAAT 2742
741 lduleuaurgvalysleuaurghthlyleuylasplalalyrvallyseu 755
2743 AACTTCATAGTTTAGCTACTGAGAAACATTTATACACATAATTC 2792
756 valglugllythlyleuaurghglulysasnleuasnghileuall 772
2793 ATTAAGTAAAGATGATATATACAGAAAGTCCCTGAAACGCTAAGA 2842
772 aspgllyserglulysalagluallaleuphspasplleuylvalleu 789

2843 GGAAGCAAAACTTCGAAATCAAAAGTATGATGAAATGGAATTCGA 2892
789 lahlsvalglmetasplargaspalvalglulaleuengllypofthr 805
2893 CAATAATGAAAGATGAGAAAGATGCTGAGCTGAAATTCCTCTTGG 2942
806 aspthrilleasprghlsvalhlsghllyleugllyleuvallyseglu 822
2943ATTGAGAAAGTTTACTTACTTGAAGAAAGT 2974
822 lgluaspelugllyrvalaleuaspserleuylarggllyvallyserl 839
2975 TAGGATCTTGA.....AACGTT 2994
839 englulugllyleugllyleuasnphelugllyllyasrgh 855
2995 CTAAAGATGATCTGAGAGCTATTCATTTATACGATGAGATGCT 3044
856 leuilevalglulvalasplaleuaurghglulhlsargmetleuasn 872
3045 ..ATTCAAACTGTGTATGATATTAAGGACCAACAAAGAGATGATGA 3091
872 uaspmet.....serlealagluvalargtrph 882
3092 TTTCTTACGTAATGAGGAAACATTTACGATTCGAAATGGA..... 3137
882 lsalasnlaargglugllyllysalysrserleuylleugllyargp 898
3138AAAGATGAAAGGTGAGAGAAATTCAGAAATGATTAATCTTG 3179
899 glulysrserglulugllyleuvalleuallaglugllyllyseglulne 915
3180 ATTAAGAAAGAA.....TTTACAGTTTGTGAAATGAAATCAATC 3220
915 ullevalglulysleuylleugllyser..... 925
3221 TTTTACACGAAACAAATTTATGATGATTTAAGTCAAAAGGAA 3270
926leuaspserleuylserglulysglulserleuengl 938
3271 ACATCAACGATATGATGATGATGATGATGATGATGATGATGATGAT 3317
939 glulysalaleuylsgllyllysalysrserleuylleugllylly 955
3318ATTTCATTAATAAAAGAAAGATGAAAGCCCAAGCTTTTGA 3361
955 agluarglys.....arglupheugllylly 965
3362 CAAGTAAAGAAATGAAAGCTGATTTCAAGTACGCAACAAACAAACG 3411
965 euaspalaleugllyargleu..... 971
3412 TTGCAGATATTAATCGTTAATAGATGATTTACAGCAATTTAAGAA 3461
972asmetlysllelysglytyrleuaspserlysalysnlylly 986
3462 GTGGTGTATTTGAAGCTTAAAGGCTTC.....GAAGAAATGGAACAA 3505
986 sleuylsgllyleugllyllyarghlsvalleuylserghlsrserleu 1003
3506 AATAAAGAACTGAA.....TTGAACAAAGCTCAAAATCTG 3543
1003 ercysmetalalysglulnargllyseralaglualeuasnlyserlys 1019
3544 AACTAAAGAGAGAGCTGATTCAGCAAAATGATTAAGAAAGAAAG 3593
1020 gluleuengllyllyllyllyllyllyllyllyllyllyllyllylly 1036
3594 AGAAACTGCTGCTCAATTAATGAGGAAACAAACCTTGAAGGAGACT 3643
1036 ulyslyrarglyllyllyllyllyllyllyllyllyllyllyllylly 1053
3644 AGAATTAATGATTTGAATACAACTTCAGCATTTGAATTCGATTA 3693

[illegible]

```

DEFINITION S. cerevisiae chromosome XIV reading frame ORF YNL250w.
ACCESSION Z71526 YJ13139
VERSION Z71526.1 GI:1302292
KEYWORDS
ORGANISM Saccharomyces cerevisiae.
SOURCE Saccharomyces cerevisiae
AUTHORS Eukaryota: Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
REFERENCE Saccharomycetales; Saccharomycetaceae; Saccharomyces.
JOURNAL 1 (bases 1 to 5078)
AUTHORS Sen-Gupta, M., Gueldeher, U., Behnhauer, J., Fiedler, T. and
Hegemann, J.H.
REFERENCE 2 (bases 1 to 5078)
JOURNAL MIPS.
AUTHORS
TITLE Direct Submission
JOURNAL Submitted (29-APR-1996) Data collected by MIPS on behalf of the
AUTHORS European yeast chromosome XIV sequencing project. MIPS at the
JOURNAL Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152
JOURNAL Martinsried, FRG; E-mail: Mews@mips.embnec.org
FEATURES
Source
1. 5078
Location/Qualifiers
1. 5078
/organism="Saccharomyces cerevisiae"
/db_xref="taxon:4932"
/chromosome="XIV"
1095..5033
/gene="RAD50"
1095..5033
/gene="RAD50"
/feature="RAD50"
/feature="ORF YNL250w"
/codon_start=1
/protein_id="CA96157.1"
/db_xref="GI:1302293"
/db_xref="Swiss-Prot:P12753"
/translation="MSA1YKLSIGISFSPDSNDRETIIEGKPLILLVGMGSGKTTII
ECLVATATGDLPEPMSKGVFIHDPKIRGSELDIAOVLAFTSAGLMNIYATRIOLML
KTTTTRKTLEGGALVAINNSGDRSTLSFSLDAOVLPGVGLKALIEVYIPCHODEB
SLMPLSPSNKKKKKKFDELFPMKRTKLDNLIKSKKMSVDIKLLKOSVHLKDKDR
SKAMKINHOLQRTIDYIDNEVSFESQLNEITFSKRLKRSNDPDKRILSKYENLKKR
TKLSIDQVKRLSNSIDLILDSKPDQLNLANFSKVLADKNNQDLNETDISLKDRQ
SSLOSLSNLSLRQGELEAGKEIYEKRNHLSKEAFQKFGSLNIENSDMAQVNH
EMSGFKAFISODLTDTIDTIDOFADKIOLEKTNINSLDIKSITVDSOMLEYNKKDKRSKLIHD
SEELALKEKSFKSLSTODSLNHELLENKTYKEKLSQMSSESIIPKLNKITEENKEMMI
ILENQIKERFQDRIMKTNOQADLYKLLIKKISNTKTLDELOKITEKONSRIPOVPP
LTQFQRADEMDQKLEINNOKKATANKKMHEDLRYTALYNLTIEKLDONOKR
SKEVYILKLESENLEDECTIDEYNVLEETELSTYTALENLMHOTTLEFNKALEIAE
RDSQCYLCSRFENESKSLQELKRTKTAPEKTLDTVONKEYLHSHRLKEKIH
ITLUSINEKINDOSCKLEKAEETKSKSLDELEVSTKLDEKEELAESIRPLEKSI
FTYLSEKLEKIDENSSTKITEELSIYNTSEDSIGTVDELDDQQRKMSDIRELRTISD
LQMKDEKVRNSMNLNLIKELTVEIESLSLQKONIDISTSKENINDIDSPVK
ELEKRTISLAKRKDEAOSVLDKVAENEDIOVRNOKTVADINRLIDRFQTYINVQDE
EAKFELISLAKRKDEAOSVLDKVAENEDIOVRNOKTVADINRLIDRFQTYINVQDE
LIEKLSOLOHETSEISRLDVONAAEERDKYEEBSLRPTREKLSNAGLGMKQD
ONIDSLTLOHARTDYKIDIEKHYKKEWELQYRSDPIDIVYSKALDSIAIKYHGLMKK
ODINRATIDELMKRTYSGTDIDTITIRSDDEVSSTYVKGKSYNRYVMYKODVLMRGCG
SAGKRVATSLIRLALSTETFGANGCVYALDEPTTNLDBENESLAKSLHNTINNRHQ
KNFQVITITHEKRLVGHMNAAFPTDFPKAKDRDROKSQLEWVDINRVTY"
BASE COUNT 1947 a 811 c 965 g 1355 t
ORIGIN
Alignment_scores:
Quality: 1477.50 Length: 1404
Ratio: 1.666 Gaps: 41
Percent Similarity: 63.177 Percent Identity: 29.701
Alignment_block:
US-09-538-396-2 x SCYNL250W
Align seg 1/1 to: SCYNL250W from: 1 to: 5078
1 MetSerThValAlaPlysmetLeuIleYsgLIleArgSerPheaspPr 17
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1095 ATGAGGCGCATCATCAATATCATATTCACAGGGGCAATACGGCTCTTTTACAC 1144

```

17 oaspsnlysasnvalllethrPhehlyserProleuthrleuileValg 34
1145 CAATGATGAGAACTATTGAAATTTGGCAAGCTCTGACTTTAAATGCG 1194
34 lProasnlyValaglylserThrThrleleleuylserleuylser 50
1195 GCATGATGCTCAGGTAAAGACATCATGCAATGTTAAAGTACGCT 1244
51 CysThrlygluleuProAsnserArgserlyHsthPheValH 67
1245 ACCAGAGGATGATCCGCCCAACAGGAGGAGA...GTATTCATTC 1291
67 sasProlyValAlaglyglulthrlysglylnleleuVal 84
1292 TGACCCGAGATGATGCTGAAAAGGACATTAGCTCAGGTCAAACCTG 1341
84 rPheThrAlaAlaglylserValAlCyslleArgserPheGln 100
1342 CGTTACGAGTCCCAATGCAATATCATGTCACCAAAATATTCAG 1391
101 leuthrGlnlyAlaSerlysmetglupheValAlleleuValle 117
1392 TTGCTAATGAAAAGACTACTACTACTATTAGACTTGAAGCGCAGTT 1441
117 uGlnThrleasnProHsthThrlyglulysValCysleuSerTyArg 134
1442 GGTGCTATTAAT...AATAGGTGTGACCGCAGTACTTGTCCACCCGG 1488
134 yAlaAspMetAspArgglulProAlaLeuMetGlyAlaSerlyAla 150
1489 CTCTGCAATTAACGACACAGTACCCCTGATCTGAGGTCCAAAGCG 1538
151 ValLeuGlnAsnVallePheValHIsGlnAspGluSerAsnTrpPro 167
1539 ATCTTGAGATGATGATATTCGTGACCAAGAGACAGCGCTGGCCACT 1588
167 uGlnAspProSerThrleuylserlyPheAspAllePheSerAlar 184
1589 GAGTGAAACATCCAACTTAAAGAAAGTTGATGAATTTTCCAAAGCG 1638
184 hrArgThrlyAlaAlaGlnValleleuylserlyHsthlyAspGln 200
1639 TGAATATTCAGAAAGCTCTAGCAATTTAAATCCATAAGAAAGACATG 1688
201 MetGlnGlnleuylserThrPheArgleuylserleuGlnleuGln 217
1689 TCCGTGATATCAAGTTATTAAAGCAATCAGTGAACACTTAAAGTTG.. 1736
217 llyAspGlnAlaHsthlyLeuArgGlnleuGlnleuGlnleuGln 234
1737GATTAAGACA 1746
234 ySerAspAlaSerlyserGlnMetGlnleuylserlyGlnleuylser 250
1747 GATCGAAACAGTAAAGCTCAATATCCACCAATTCAGAACTTAAGATGAT 1796
251 GlyThrGlnArgGlnleuGlnMetGlnThrSerleuAspGlnleuArg 267
1797 CAGTAAATGAGAGAGTCTGTGAATTCATCATTCATGAATGAATC.. 1844
267 gArgleuGlnGlnleuAspIleuylserAlaThrGlnArgSerThrleu 284
1845ACTGAGAAATCCGATTAAC 1863
284 eu.....ThrGlnGlnHsthlyserleuAlaAlaLeuSerGlnGln 297
1864 TGTTTAATCTATCAAGATTTCCAAAATACTATCCAAAGTAAAGAAC 1913
298 AsnGluAspThrAspGlnleuMetGlnThrGlnThrlyAspPheGln 314
1914 TTGAAAAATACAAACTCTCAATTAAGTATGATCAAGTAAAGATTTGCAAA 1963

314 uArgleuAlaLeuGln.....T 321
1964 CTCAATGATATCTGATTTATCGAAGCAGATTTGCCAAACTATTTGG 2013
321 hrlyleuSerlyleuValArgAsp.....MetAspArg 332
2014 CAATCTCTTAAGTCTGATGATAGATTAAGATTAAGTATGATGATTTG 2063
333 GluAlaSerlyserSerValleuSerlyGlnAsnSer.....Glnle 347
2064 GAAACCGATTTTCAAGTTGAAAGACCGCATCTCATCTGCAAGACCT 2113
347 uThrHsthGlnleuGlyleuGlnleuAlaAlaAspAlaHsthleuThr 364
2114 TTCTATTCCTCATCAGGCGGCGAGGTAGTAAAGT..... 2153
364 ellyHsthGlnArgAspSerAspIleuAsnleuSerlyThrlyHsthAsn 380
2154GTTAAGAGACATACGAAAGAAACAGAAATCACCTTTCATCG 2195
381 leuGlyProValProGlnHsthProPheThrAsnAspValAlaMetAsnle 397
2196 CTAAAGAAAGCTTCCACATTAAGTTC.....CAAGGCT 2230
397 uThrAsnArgIleuValAlaArgleuSerSerleuGlnleuAsnAspLeuVal 414
2231 TTCAAAATATTGAAACGATATGCTCAAGTAAATTCAGAAATGTCAC 2280
414 sp...LyslylserAsnGlnAspGlnleuAspValleuThrlyHsth 429
2281 AGTTCAAGCTTTCATCTCAGATTTAAGTATTAAGTATTAAGTATTA 2330
430 TyrleuylserAlaAlaArgTyserGlnleuValAspGlnleuGlnle 446
2331 GCAAGAGATTTCACTAAAGAAACAAATTTTCCGACTGTTAAATAC 2380
446 llyHsthleuSerlyleuArgGlnleuArgGlnleuArgGlnleuArg 463
2381 AATTAACCGTATTCACAGATTTGAAATTAATTAATTAATTAATTAAT 2430
463 ys.....GluArgAspAlaAlaGlnleuValleuSerlyPhe 475
2431 AATTATTCATGACTCGAAGACTTCTGAGAAATTAAGAAATGATTA 2480
476 AsnleuSerArgleuAspGlnleuArgGlnleuGlnleuGlnleu 491
2481 AGTTTGTCTCTCAAGTATCCCAATCAATCAATCAATCAATCAATCA 2530
491 lGluArglyThrleuAlaLeuGlnleuArgAsp.....T 503
2531 ATACAAAGAAAGCTCGAATCTTGGGAAAGTGAATATTATTCAAAT 2580
503 yAspSerleuSerleuGlnleuArgGlnleuValleuSerleuGlnleu 519
2581 TGAACCGAAATTAAGAAAGAAATTAATTAATTAATTAATTAATTA 2630
520 llyleuValleuLeuArgGlnleuArgIleuAsnArgAsnAlaAs 536
2631 CAATTAAGAAATTTCAAGATCCATTAATGAAGAAACCAACCAAGCG 2680
536 pGluArgValleuGlnleuGlnleuArgGlnleuArgGlnleuArg 553
2681 CTATTAAGCGCAATCAAGTATTAAGTATTAATTAATTAATTAATTA 2730
553 sPlyleuAsnGlnleuValleuAsn.....GlnleuAspIleuValleu 567
2731 ACAGATTACAAAGATTAAGGAAATTAAGAAATTAAGTATTAAGTAA 2780
568 llyValleuArgGlnleuArgAsnProPheGlnleu...AspMetlyAsp 583
2781 CAAGTATTTCCCTCAACCGAGAAATTTCAAGAGCGATTAAGTAAAT 2830
583 uleuAsnGlnAlaPheTrpProValAspLyGlnleuArgGlnleuArg 600

[illegible]

1162 1aleuMetarphenisserMetlyMetGluGluIleasnlyIlelle 1162
 1163 1lysGluLeuTrpGlnGlnThrTyraGlyGlnaspIleasPTyIlele 1179
 4554 GATGATATATGAGCGTACGATGATGATAAAGCAAGATATAACAGATAAAT 4553
 1179 rIleasnserasp.....SergIuGlyAlaGlyThrargSerTyserT 1194
 4604 AATTGATCAGATGAGTACGATGATGATGATGATGATGATGATGATGAT 4653
 1194 rYraGlyValMetGlnThrGlyaspAlaGluLeuGluMetargGlyarg 1210
 4654 ATCGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4703
 1211 CysSerAlaGlyGlnGlnValLeuAlaSerLeuIleIleArgLeuAla 1227
 4704 TGTTCGGGGGCAAAAGTCTTCATCGATGATGATGATGATGATGATGAT 4753
 1227 uAlaGluThrPheCysLeuasnCysGlyIleLeuAlaLeuaspGluProT 1244
 4754 ATCTAAACATTTGGCGCAAAATTTGGCGTATTCACATGACGACCACTA 4803
 1244 hrThAsnLeuaspGlyProasnAlaGluSerLeuAlaAlaAlaLeu 1260
 4804 CCACCAATTTAGATGAGAAACATGCAAAAGTTAGCAAAATCTTTACAC 4853
 1261 ArgIleMetGluAlaArgGlyGlnGlnGlnGlnGlnGlnGlnGlnGln 1277
 4854 AATAAATAATATGATGAGACATCAAAAGAAATTTCCACATGATGATGAT 4903
 1277 ethrIAspGluArgPheAlaIleLeuIleGlyGlnGlnGlnGlnGlnGln 1294
 4904 CACTATATGATGAGAAATTTTGGGTCATGATGATGATGATGATGATGAT 4953
 1294 IuLyTy 1310
 4954 ATCACTCTCTCAAGTCAAGAGATGATGATGATGATGATGATGATGATGAT 5003
 1311 SerGlnGluIle 1314
 5004 TGGGTCGATATT 5015
 seq_name: gb_p13:SCCHXIVL
 seq_documentation_block:
 LOCUS SCCHXIVL 33016 bp DNA PLN 30-JUL-1997
 DEFINITION S.cerevisiae DNA region from chromosome XIV, left arm.
 ACCESSION X96722
 VERSION X96722.1 GI:1255956
 KEYWORDS ATX1 gene; MPN3 gene; NRD1 gene; ORC5 gene; RAD50 gene; RPA49
 gene; SIP3 gene; su11 gene.
 SOURCE Saccharomyces cerevisiae.
 ORGANISM Saccharomyces cerevisiae.
 REFERENCE 1 (bases 1 to 33016)
 AUTHORS Ben-dupa,M., Guldener,U., Beinhauer,J., Fiedler,T. and Hegemann,J.H.
 TITLE Sequence analysis of the 33 kb long region between ORC5 and SU11
 JOURNAL Yeast 13 (9), 849-860 (1997)
 MEDLINE 97377992
 REFERENCE 2 (bases 1 to 33016)
 AUTHORS Hegemann,J.H.
 TITLE Direct Submission
 JOURNAL Submitted (21-MAR-1996) J.H. Hegemann, Institute of Micro &
 Molecular Biology, Justus-Liebig-Univ. Giessen, Frankfurt Str.
 107, D- 35392 Giessen, FRG
 FEATURES
 source I. .33016
 Location/Qualifiers

/organism="Saccharomyces cerevisiae"
 /strain="S288c"
 /sub_strain="FY1679"
 /db_xref="taxon:4932"
 /chromosome="XIV"
 /clone="cosmid 14-4c pou"
 /map="132kb, left arm"
 223..1662
 /gene="ORC5"
 223..1662
 /gene="ORC5"
 /note="ORF N0834"
 /codon_start=1
 /protein_id="CAA65483.1"
 /db_xref="GI:1255957"
 /db_xref="SWISS-PROT:P50874"
 /translation="MNTPEVAFREYQTCCLASYSADPDTPSNLLIQYSGTGT
 YTKKYNPNPNHAWLEPEVELSNKPLIDALAPVQYKFLTYNIPPTDDPQY
 EEPFLYKTHNIFVQESQERTCLFLIDGSDQDDAALFNKYIKLNEELPDS
 KINKFLTYMLETSLQRYSTHCLPVMFPRIVNDEYSTLVNSRGELMDESLKRR
 IIEQITDCTDDQFQNVANFHLIVQAFSYGNDIFALNDLIDRWKPYRSITKE
 NIEPFLALYSAIKLEFLSTDNLSNQGSAATTTNDLNSQTYDLSTISLYLLA
 SYIOSYEPYDASIFERTKIIOGRAAYRRKKKENVPRYLOPSFAIERLAIQOA
 IPIQGAESGSSALREESIMKANTIEVPQNLSELHLKLATNMKNIDYSPYRM
 KANVPETIKREISSVFNISDYFSQIHE"
 complement(1392..2578)
 /note="ORF N0838"
 /codon_start=1
 /protein_id="CAA65484.1"
 /db_xref="GI:1255958"
 /db_xref="SWISS-PROT:P53846"
 /translation="MVRNRFIRKRNKLFNSNLSYKSKWKYKITQIKMDPNLLN
 LEEQYQEGLEGENIKQSLFEGQYQVQVQFORTELLQMEGLQVDESGLSP
 TLEKNHITRTLMGLMKNNDSEVSEFEFVLLIKKKRFTITLIRLVKDKRFTFY
 TFEYEDVSRAIGEIRGEFVNEEDIAKKTKQNOAQSW"
 complement(2767..2988)
 /gene="ATX1"
 complement(2767..2988)
 /note="ORF N0840"
 /codon_start=1
 /protein_id="CAA65485.1"
 /db_xref="GI:1255959"
 /db_xref="SWISS-PROT:P38636"
 /translation="MAEIKHYOFNVVWTCGCGGAVNKVLTJLKEPDVSKIDISLEKOL
 VDYVTLTPYDFILEKIKTKGKGVNSKOL"
 complement(3323..5496)
 /note="ORF N0842"
 /codon_start=1
 /protein_id="CAA65486.1"
 /db_xref="GI:1255960"
 /db_xref="SWISS-PROT:P53847"
 /translation="MESLFPNKGELIRELKDPLILKNSRNSGSELDESSDLOR
 EAILANELINDIKFTPLNLIKKYNLTNLEBENCYSGSLRKKRNNAATLKSP
 NFQOISITVYDLELVSTLYLITNGFKITENSIOPTVEMGDKVHITDTFM
 DPAVOQYFPGSDLDQAMFLDWTASQSOQVAKLTITKEKYNLSRYLSMKNSIF
 ISGREISYENKNTLVSKSSSHQCHSVYLSFEVACDFMDLAFDRKTLSTEL
 GPLENTEFTKFNKNAISILLESIDSPKNTVYNSFKLTPVAKSEVTNTHSGKEIO
 DILMNKQLYNLLDKYLSHISIRISIFEDPKSKMONTLEVLTSSNNTMSEKIGK
 NDSVONKEKLHNAVSQDDDMNREVEDDDANGDELDVNVIDDEBETKQEKREBE
 ENNAWDEAMALIDENIDASLENKEHLKADVGSGLDHLLEVQLPEPLAISQNS
 SPADSHVDQYFAKYNLQTSYAMKQTFANSHMCOLVYDMYLLERDKLRIKEL
 TRNLEFKLNNKYRIVCOLIRHOLTEPRENERPNSMATAIEKLLPYLKEIVPLRIKEL
 RGEYSRYLSLEFLNDCVTEKILKMOIIEVNSNSELGELVSLVNNNDIOLAKE
 PSYKMKREKATMGKFLPLILKEIMEFVNGDFYLFATDILQIMTELLFADPTLRNA
 IDDIYETRGTRALDD"
 complement(5754..9443)
 /gene="SIP3"
 complement(5754..9443)
 /gene="SIP3"
 /note="ORF N0844"
 /codon_start=1
 /protein_id="CAA65487.1"


```

20927 CTCGTGATTTAGCCACAGACAGTACCCCTGATCTGGAGTTCACAAAGCG 20976
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
151 VALLEUUAASNVALLIephEVALHISGLNASPGLUSeRsnTrpProLe 167
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
20977 ATCTTGAAATACGTGATATCTGTGCACCAAGAGACAGCTGTGCACCT 21026
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
167 UGLNASPRosERThrLeuLYsLYsLYsLYsPhEASpRIePhEsERAlAt 184
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
21077 GAGTGAACCATCCAACTTTAAAAAGAGTTGTATGATAAATTTTCCAAAGCA 21076
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
184 hrArgTYThrLYsAlaLeuGLUValLIeLYsLYsLeuHISLYsASpGln 200
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
21077 TGAATTTACAAAGCGCTCTAGCAATTTAAAAATTCATAAGAAAGCAATG 21126
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
201 MetGLInLIeLYsThrPhEArgLeuLYsLeuGLUASnLeuGLnThrVa 217
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
21127 TCCCGATATCAAGTATTTAAAGCAATCAGTAGAACACTTAAGTGTG.. 21174
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
217 LLYsASpGlnAlaHISLYsLeuArgGLUASnLIeAlaGLNASpGlnGLU 234
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
21175 ..:::.....GATAAAGACA 21184
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
234 yASERASpAlaSERLYsSERGLnMetGLnLeuLYsGLUlySILECYs 250
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
21185 GATCGAAAGCAATGAGCTCAATATCCAAATTCGAAACTTAAGATAGAT 21234
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
251 GLYThrGLUArgGLUleuLeuGLnMetGLnThrSERLeuASpGLUleuAr 267
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
21235 CAGTATATGAAAGAGTGTCTGAATTTGAATCTCAATTAAGTAAGAAATC.. 21282
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
267 gArgLeuGLnGLYGLnLIeASpRIeLYsAlaThrGLUArgSERThrLeuL 284
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
21283 ..:::.....ACTGGAATTCGATTAAC 21301
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
284 eu.....ThrGLnGLnHISGLYsLYsLeuAlaLIeLeuSERGLnGLU 297
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
21302 TGTTTAAATCTATATCAAGATTTCCAAAAATACTATCCAAAGTAGAGAAC 21351
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
298 ASnGLUASpThrASpGLUleuLeuMetGLnUTRGJnThrLYsPhEGLnGL 314
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
21352 TTGAAATAATCAAAACTCTCAATTAGTGATCAAGTGAAGAAAGATGTCAAA 21401
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
314 uArgLIeAlaLeuLeuGLU.....T 321
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
21402 CTCATTTGATATTCGATTTATCGAACCGACATTTGCAAAACTATATG 21451
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
321 hrLYsLIeSERLYsLeuValArgASp.....MetASpASp 332
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
21452 CAAACTTCTCTAAAGCTTGATGATTAAGAAATATCAGTTGAGAGATTG 21501
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
333 GLUAlaSERTYrSERSerValLeuSERLYsGLNASnSER.....GLUde 347
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
21502 GAACCGATATTTCAGATTGGAAGACCGTCACTCTCTGCAGAGCGCT 21551
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
347 uThrHISGLUleuGLYsLYsLeuGLnAlaGLUAlaASpAlaHISLeuThm 364
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
21552 TTCTAATTCCTCCATCAGCGCGGAGGTGAGTTAGAACT..... 21591
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
364 eLYsHISGLUArgASpSERASpRIeLYsASnLIeCYsThrLYsHISASn 380
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
21592 ..:::.....GCTAAGAGACATCGAAAGAACAGAAATCACCTTTTCATCG 21633
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
381 LeuGLYProVALProGLUHISeRPhEThrASnASpVALAlaMetASnLe 397
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
21634 CTAAAGAGAGCGCTTCCACATAGATTC.....CAAGCGCT 21668
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
397 uThrASnArgLIeLYsAlaArgLeuSERSerLeuGLNASnASpLeuLeuA 414
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
21669 TTCAATATATGAAACAGTATATGCTCAATAATTCACGAAATGTCA 21718
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
414 sp...LYsLYsLYsSERASnGLUASpGlnLeuASpAlaLeuTrpLYsHIS 429
      ::::::::::: ::::::::::: ::::::::::: :::::::::::

```

```

21719 AGTTCAGCGTTTCATCTCTCAGATTTAACTGACACTAATCAATTT 21768
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
430 TYrLeuLYsLIeASnAlaArgTYrSERGLUValASpGLYGLnLIeGLnSe 446
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
21769 GCAAAGAGATTTCACTTAAGAGAACAAATTTATCCGACTGTGATAAATC 21818
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
446 rLYsLIeGLUSeRMetSERGLYLIeLeuArgArgArgLYsASpLYsGLU 463
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
21819 AATTAACCGTCGATTCACAGATTTTGGAAATTTAATAAAGGACAAATCTA 21868
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
463 yS.....GLUArgASpAlaAlaGLUValGLUleuSERLYsPhe 475
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
21869 AATTAATCATGACTGTGAAGAGCGTTCGAGAAATTTGAATCAATTTAA 21918
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
476 ASnLeuSERArgLIeASpGLUArgGLUArgHISMetGLn..LIeGLUVA 491
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
21919 AGTTTGTCTACACAGATAGCTCAATCATGTGAACTGGAACCTTTAAAC 21968
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
491 LGLUArgTYrThrLeuAlaLeuGLYGLUArgASp.....T 503
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
21969 ATTCAAAGAAAGCGCAATCTTGGAAGTGAAGATTTATTAACCAAAAT 22018
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
503 yRASpSERLIeIleSERGLnLYsArgTYrGLUValTYrSERLeuGLnGL 519
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
22019 TGAAACGAGAAATTAGAGAGAGAAATTAATAGATGATTTCTTAGAGAT 22068
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
520 LYsLIeLYsValLeuLeuArgGLUlySASpRIeIleASnArgASnAlaAS 536
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
22069 CAATATGAGAAATTTCAAGATCGCATTAAGAAAGCAACCAAGCGGA 22118
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
536 pGLUArgVALLYsLeuGLYLeuLYsLYsASpAlaLeuGLUSeRLYsA 553
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
22119 CTTATACGCCCAACTAGAGCTAATAAGAGATCAATTAATCTAATAATAG 22168
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
553 sPLYsLeuASnGLUleuValASn.....GLUHISeLYsASpLYsILECYs 567
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
22169 ACGAGTTACAAAGATTTACGGAAGAAATTCACAGAAATGTCTGAAATAG 22218
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
568 LYsVALLeuArgGLYArgASnProPhEGLUlyS...ASpMetLYsGL 583
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
22219 CAAGATTTTCCCTCACCCAGAGAAATTTCAAGAGCGATTTTGAATGGA 22268
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
583 uLIeASnGLnAlaPhETrPProVALASpLYsGLUtyrASnGLUleuAys 600
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
22269 TTTTCAGAAATTAATTCATCAATATGCAAAAGAAATATAGCGATCAACA 22318
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
600 eLYsSERGLnGLUAlaGLUGLnGLUleuLYsPhETrGLnSERLYsVAL 616
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
22319 AANAAGTCACGAAATTTGATAG.....AGATAT 22347
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
617 ThrASpAlaArgGLUleuThrLYsLeuArgArgASpMetASpAlaLY 633
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
22348 ACAAAAGCTTTGTACAAATTTGAACACCAATTTGAAGAGATTTACAGATTA 22397
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
633 sArgArgPhELeuASpSERSerLYsLeuGLnSERIle.....LeuGLnI 647
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
22398 TCAGAGTGTGAAGAAAGAAAGTATACACTACTAAGCGAAACTTAACAG 22447
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
647 IeSERAlaASnVALASpMetPhEProLYsVALLeuGLnASpAlaMetASn 663
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
22448 AGGATGTACTATTGATGATTAATGATGCTCTGGAAGAA.....ACG 22491
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
664 LYsArgASpGLUleuLYsArgLeuGLUASnPhEAlaASnGLYMetArgGL 680
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
22492 GAGCTCTCTAATAAGACGCGCTGGAAGAACTTGAAA.....ATGCATCA 22535
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
680 uMetLeuAlaProPhEGLU.....HISLeuAlaArgLYsASnH 693
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
22536 AACTACTTTAGAAATTAATAAGAAACATTAAGAGATCGCTGAACGTGACA 22585
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
693 IValLYsProCYsCYsGLUArgAlaPhETrProASpGLUleuASpGLU 709
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
22586 GCTGCTGTATTGTCTCTAGAAAGTT.....GAAATGAATGCG 22626
      ::::::::::: ::::::::::: ::::::::::: :::::::::::

```



```
24292 AATATATATATATGAGAGACATCAAAAGATTTCAGCTAATCGTGT 24341
1277 ePhrHisAspGluArgPheAlaHisLeuIleGlyGlnArgGlnLeuAlaG 1294
24342 CACTCATGATGAGAAATTTTGGTCATATGAATGCGCGCATTTACTG 24391
1294 LuLysTyrTyrArgValSerLysAspGluAsnGlnHisSerIleIleGlu 1310
24392 ATCACTTCTTCAAGTCAAGAGAGATGATAGGCCAAAAATCACAAATTGAG 24441
1311 SerGlnGluIle 1314
24442 TGGGTGATATT 24453
```


Alignment block:

US-09-538-396-2 x AM275863/rev

Align seg 1/1 to reverse of: AM275863 from: 1 to: 596

```

1140 TyrThrAlaLeuAspLysAlaLeuMetArgPheHisSerMetLysMetG1 1156
|||||
596 TATAAGCTTCGGACCAACATATGAAATTTTCACATGAAATGCA 547
1156 uGluIleAsnLysIleIleLysGluLeuTrpGlnGlnThrTyrArgLysG 1173
|||||
546 AGAATCAATAAATAATACGCTGCGCAAGTACCTATCGCGGAC 497
1173 LnaAspIleAspTyrIleSerIleAsnSerAspSerGlu..... 1185
|||||
496 AAGATATTGAATACATAGAAATACGGTCTGATGCGGATGAAATATATCA 447
|||||
1186 ...GlyAlaGlyThrArgSerTyrSerTyrArgValValMetGlnThrG1 1201
|||||
446 GTTCTGTGATAAAGCGCAATATACACGAGTGTGATCGTCAAGG 397
|||||
1201 YAspAlaGluLeuGluMetArgLysArgCysSerAlaGlyGlnLysValL 1218
|||||
396 AGACACACCTTGATGATGCGAGACGATGAGTCTGACAAAGCTAT 347
|||||
1218 euAlaSerLeuIleIleArgLeuAlaLeuAlaGluThrPheCysLeuAsn 1234
|||||
346 TAGCTTCATCATCATCGCTGCGCTGCGTGAAGCGTCTGCTCAAC 297
|||||
1235 CysGlyIleLeuAlaLeuAspGluProThrThrAsnLeuAspGlyProAs 1251
|||||
296 TGTGCAATCATTTGCTGTGATGAGCCACACAAATCTTGACCGAAGAAA 247
|||||
1251 nAlaGluSerLeuAlaAlaAlaLeuLeuArgIleMetGluAlaArgLysG 1268
|||||
246 CATGCAATCTCTTGACATGCTCGTGTGAGATATATAAAGTCCGTCAC 197
|||||
1268 IyGlnGluAsnPheGlnLeuIleValIleThrHisAspGluArgPheAla 1284
|||||
196 AGCAGCGTAACCTCCAGCTTCGTGTAATCACAATGATGAAATTTTGG 147
|||||
1285 HisIleuIleGlyLysArgGlnLeuAlaGluLysTyrTyrArgValSerL 1301
|||||
146 GAGCTTTTGAAGACGTTCTGATATATGAGGAGAAATCTTACAGATTAATA 97
|||||
1301 sAspGluAsnGlnHisSerIleIle 1309
|||||
96 GAACATCGATCATCGTCTCAGAGATT 72

```

seq_name: gb_gss23:AZ216324

seq_documentation_block:

```

LOCUS      AZ216324      597 bp      DNA      09-JUN-2000
DEFINITION Sheared DNA-80B5.TF.Sheared DNA Trypanosoma brucei genomic clone
ACCESSION  AZ216324
VERSION    AZ216324.1 GI:8434124
KEYWORDS   GSS.
SOURCE     Trypanosoma brucei.
ORGANISM   Trypanosoma brucei.
            Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma.

```

REFERENCE
AUTHORS
1 (bases 1 to 597)
El-Sayed, N., Zhao, S., Zhao, H., Gill, S., Sub, E., Malek, J., Fujii, C.,
Gerrard, C., Leech, V., de Jong, P., Ullu, E., Melville, S., Donelson, J.,
Praser, C. and Adams, M.

TITLE
Determination of clone end sequences from Trypanosoma brucei GUTat
10.1 sheared DNA library

JOURNAL
COMMENT
Unpublished (1999)
Contact: Najib M. El-Sayed
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA

FEATURES
source

Tel: 301 838 0200
Fax: 301 838 0208
Email: nelsayed@tigr.org
Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
DNA library constructed at TIGR. Clones will be available for
distribution through Research Genetics, Alabama, USA. Sheared DNA
end sequences search page: <http://www.tigr.org/cdb/mdb/cdb/>.
Seq primer: M13-Forward
Class: shotgun.

Location/Qualifiers
1..597
/organism="Trypanosoma brucei"
/strain="TREU927/4 GUTat 10.1"
/db_xref="taxon:5691"
/clone="Sheared DNA-80B5"
/note="Vector: pUC18; Site_1: SmaI; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically
sheared to give a tight size distribution (approx 2 kb).
The v + l method used for the library construction is
described in detail in Smith, H.O. and Venter, J.C.
(Making small insert libraries for whole genome shotgun
sequencing projects. In Genome Sequencing: A Practical
Approach, eds. M. Vaudin and B. Barrel, Oxford University
Press, 1999)."

BASE COUNT 160 a 134 c 156 g 147 t
ORIGIN

alignment_scores:

Quality: 561.00 Length: 194
Ratio: 3.380 Gaps: 2
Percent Similarity: 85.567 Percent Identity: 54.639

Alignment block:

US-09-538-396-2 x AZ216324

Align seg 1/1 to: AZ216324 from: 1 to: 597

```

1112 TYRYSASPILGLULYSARGTYTHRANGLINPHELEUGINLEULYSNTH 1128
|||||
1 TCAAAAGATATGAAAGCGCTACCGACGACCTTCTCGAAGTCCAGAC 50
|||||
1128 rThrgluMetAlaAsnLysAspLeuAspArgTyrThrAlaLeuAspL 1145
|||||
51 TACTGAATTTGCGGTTTCGATGTGAAAAATATATATCGACGCGTGAAA 100
|||||
1145 yAlaLeuMetArgPheHisSerMetLysMetGluGluIleAsnLysIle 1161
|||||
101 AGCAGCTGACAGACTACACACAGAGAAATGCGCATCAATCAATC 150
|||||
1162 IleLysGluLeuTrpGlnGlnThrTyrArgGlyGlnAspTyrIle 1178
|||||
151 TTGGCTGACCTTTGGCGACACACCTCAAGGAGAGATATGACACAT 200
|||||
1178 eSerIleAsnSerAspSerGluGlyAlaGlyThr.....ArgSerT 1192
|||||
201 TGAACCTCCGTCGAGATGATGTCACGCACTACAGCAGAGCGAGATT 250
|||||
1192 ySerTyrArgValValMetGlnThrGlyAspAlaGluLeuGluMetArg 1208
|||||
251 ACAGCTATCGTGTGATGAAAGGGGGAACGTAATAATGGCATGCGG 300
|||||
1209 GlyArgCysSerAlaGlyGlnLysValLeuAlaSerLeuIleIleArg 1225
|||||
301 GCGCGTTCGAGTCGCGGCTCAAAAGATCTGCTCTCATCGCACT 350
|||||
1225 uAlaLeuAlaGluThrPheCysLeuAsnCysGlyIleLeuAlaLeuAspG 1242
|||||
351 TCGGCTAAGGAAGCCTTTGTGTGACTGCGGAATCTTGATATGATGAG 400

```


CORRESPONDENCE ADDRESS:

ADDRESSEE: Dann, Dorfman, Herrell and Skillman
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA

COUNTRY: USA

ZIP: 19103-2307

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/16216

FILING DATE:

CLASSIFICATION:

PRIORITY APPLICATION DATA:

APPLICATION NUMBER: US 08/353,700

FILING DATE: 09-DEC-1995

ATTORNEY/AGENT INFORMATION:

NAME: Reed, Janet E.

REGISTRATION NUMBER: 36,252

TELECOMMUNICATION INFORMATION:

TELEPHONE: (215) 563-4100

TELEFAX: (215) 563-4044

INFORMATION FOR SEQ. ID NO. 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 10136 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHEICAL: NO

ANTI-SENSE: NO

PCT-US95-16216-2

alignment scores:

Quality: 439.50

Ratio: 0.668

Percent Similarity: 52.389

Percent Identity: 19.904

Length: 1256

Gaps: 49

Percent Identity: 19.904

alignment block:

US-09-538-396-2 x PCT-US95-16216-2 ..

Align seg 1/1 to: PCT-US95-16216-2 from: 1 to: 10136

```
51 CysThrGluGluLeuProPheSerSerSerGlyHisThrPheValH 67
|||||
5619 TGCCTGGGAGACAGTCCCGCATCCAAAT.....TA 5650
67 SASPrOlyValAlaGlyGluThrGluThrLysGlyGlnIleLysLeuA 84
|||||
5651 TGAGCGCT.....CCAGGGAGAAATAAACCAGGCGCTTCAGAA.... 5690
84 rGpHeLysThrAlaAlaGlyLysAspValValCysIleArg.....Ser 98
|||||
5691 .....TSCATTTCGAATTGCA 5708
99 PheGlnLeuThrGlnLysAlaSerLysMetGluPhe.....Lys 111
|||||
5709 TTTCTGGTCTTAATGCTTTGTTGATGATTTCCCGGGGAAATCAGA 5758
111 AlaIleGluSerValLeuGlnThrIleAsnProIleThrGlyGluLysV 128
|||||
5759 AGATTCATATGCTTCATTCGCGGTAAGAGACATCAATGAAATTT 5808
128 ALcysLeuSerThrArgCysAlaAspMetAspArgGluIleProAlaLeu 144
|||||
5809 TGAGATTACTTCATGATGATAGAGACCGTGCACAAAAAGTTGAAAGTTTG 5858
145 MetGlyVal.....SerLysAlaValLeuGluAsnValI 156
|||||
```

```
5859 CTAATGAATGAAGAATAGACTCAAACTCCATTAGACGAGCTACA 5908
156 ePheValHisGlnAspLysSerAsnTrpProLeuGlnAsp...ProSer 172
|||||
5909 ACTATGACCAAAATTAAGACATGATGAAATTAAGTGGG 5958
172 hLeuLysLysLysPheAspAspIlePheSerAlaThrArgTyrThrLys 188
|||||
5959 AACTTAAGAAAGAAACTCAGATT.....AGTGA 5990
189 AlaLeuGluValIleLysLysLysHisLysAspGlnMetGlnIleLys 205
|||||
5991 AAATGGAAATATTTCTGTCGATCAGCAGAGCTTACCCGAGAGTAGA 6040
205 sThrPheArgLeuLysLeuGluAsnLeuGlnThrValLysAspGlnLan 222
|||||
6041 AACTTGAAGGCGCTCAATCTGATTTAGAAATGATGATGATTAATCAT 6090
222 IsLysLeuArgGluAsnIleAlaGlnAspGlnGluLysSerAspAlaSer 238
|||||
6091 CA.....CGTGAAGATATTGAGATATATGCGCCAGGTGAATGACAGC 6134
239 .....LysSerGlnMetGluGlnLeuLysG 247
6135 TGAAGAGAGAGATTTCTGATGTGAAATGACCTGATGAGATGATGATC 6184
247 uLysIleCysGlyThrGluArgGluIleLeuGlnMetGluThrSerLeuA 264
|||||
6185 GGAGAAAGCTAGCATGTGATGAGAGAGAGAGAGAGAGAGAGAGAGAG 6234
264 sPGLuLeuArgArg..... 268
6235 AGTAGTTCAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6284
269 .....LeuGlnGlyGlnIleAspIleLysAlaThrG 279
6285 CAGAGGTTATTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6334
279 uArgSerThrLeu.....LeuThrGlnGlnHisGlu 290
|||||
6335 GAGAAACAGAGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6384
290 yLeuAlaAlaLeuSerGlnGluAsnGluAspThrAspGluGluMet 306
|||||
6385 CACTGATCACTGCTGTAAGAAATGAAGAGAGAGAGAGAGAGAGAGAGAG 6434
307 GluTrpGlnThrLysPheGluGluArgIleAlaLeuLeuGluThrLysI 323
|||||
6435 TCTCATCAAACTGAGTGTCATTCATTCAGGTGCGAGAGAGAGAGAGAG 6484
323 eSerLysLeuValArgAspMetAspAspGluAlaSerTyrSerSerVal 340
|||||
6485 GAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6534
340 euserLysGlnAsnSerGluLeuThrHisGluIleGlyLysLeuGlnAla 356
|||||
6535 TGTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6584
357 GluAlaAspAlaHisLeuThrMetLysHisGluArgAspSerAspIle 373
|||||
6585 GACTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6634
373 sAsnIleCysThrLysHisAsnLeuGlyProValProGluHisProPhe 390
|||||
6635 ACAACTGATTAAGAG..... 6650
390 hAsnAspValAlaMetAsnLeuThrAsnArgIleLysAlaArgLeuSer 406
|||||
6651 ..AAGATTTCTGTCAGAGATCTGAAGCTGAGAGAGAGAGAGAGAGAG 6698
407 SerLeuGluAsnAspLeuLeuAspLysLysLysSerAsnGluAspGlnLe 423
|||||
6699 GAATCAGATTATGAAGAGCTGAATGTCTCAAGAGAGAGAGAGAGAGAG 6748
```

423 uAspValLeuThrIysHisTyrLeuLysIleAsnAlaIArgTyrSerGluVal 440
||| : : : : :
6749 GGTGGAAAGAGTGGATTCGATTCGAGCTGAGCTCAACACAGAGAGAG 6798
440 aIAspArgGlnIleGlnSerLysIleGlnSerMetSerGlyIleLeuArg 456
||| : : : : :
6799 TGCAT...CAGCTGAGAAGAGCATCGAAGAACTGAGA...GTTCCG 6839
457 ArgArgLysAspLysGluLysGluArgAspAlaIleGluValGluLeu 473
||| : : : : :
6840 ATTGAGGCCATGAAAGAACAGCTGCACATCCACAG... 6878
473 rLysPheAsnLeuSerArgIleAspGluArgGluArgHisMetGlnIleG 490
||| : : : : :
6879AACTGAAAGAACGCGACGG..... 6899
490 luvAlGluArgLysThrLeuAlaLeuGluArgAspTyrAspSerIle 506
||| : : : : :
6900GAGATGAT..... 6908
507 lIleSerGlnLysArgThrGluValTyrSerLeuGlnLysIleLysVal 523
||| : : : : :
6909TCACCTTACGATTAAGTTGAGAA 6931
523 lLeuLeuArgGluLysAspIleIleAsnArgAsnAlaAspGluArgVal 540
||| : : : : :
6932 CCTTGAAGGGAATTCAGATGTCAGAAAGAAAC...CAGGAGCTAGTGA 6978
540 yLysLeuLysLysLysAspAlaLeuGlnSerLysAspLysLeuAsn 556
||| : : : : :
6979 TTCTT.....GATGCC...GAGAAATCCAAAGCAGAAAGTACAG 7013
557 GluIleValAsnGlnHisLysAspLysIleLysLysValLeuArgGlyArg 573
||| : : : : :
7014 ACTCTA.....AAAACCAAAATAGAAAGACATGCCAGAACCT 7051
573 gaenProPheGluLysAspMetLysLysGluIleAsnGlnAlaPheThr 590
||| : : : : :
7052 GAAGATTTTGAATTAGACCTGTC..... 7076
590 roValAspLysGluTyrAsnGluLeuArgSerLysSerGlnGluAlaGlu 606
||| : : : : :
7077ACGTTAAGCTCTGAAAAAGAAATCTGACA 7106
607 GlnGluLeuLysPheThrGlnSerLysValThrAspAlaArgGluGln 623
||| : : : : :
7107 AAACAATACAGAAAAACAAGCT.....CAGTT 7135
623 uThrLysLeuArgArgAspMetAspAlaLysArgArgPheLeuAspSer 640
||| : : : : :
7136 GTGAGAACTAGACAACTTACTCTCTCATTTAAAGCTGTGTAGAAAGAA 7185
640 yLeuGlnSerIleLeuGlnIleSerAlaAsnValAspMetPheProLys 656
||| : : : : :
7186 ACGAGCAACAGAGATACAGATC..... 7208
657 ValLeuGlnAspAlaMetAsnLysArgAspGlnLysArgLeuGluAs 673
||| : : : : :
7209AAAGAAATCTTAAAACTGCACTGAGAT 7237
673 nPheAlaAsnGlnLysMetArgGluMetLeuAlaProPheGlnHisLeuAla 690
||| : : : : :
7238 GCTTCAGAACTCAGTTAAAGAGCTTAATAGAGCAGTAGACAGCTTGTGTG 7287
690 rGluAsnHisValCysProCysCysGluArgAlaPheThrProArgLys 706
||| : : : : :
7288 GTGACCAAGAAATTAATGAGGCCACAGAACAGCTAGACCCACCAATA 7337
707 GluAspGluPheValLysLysGlnArgMetGlnAsnSerSerThrAlaG 723
||| : : : : :
7338 GAGGAAGAG.....CATCAGCTGAGAAATATAGCATTCAAAGCT 7375

723 uArgSerLysAlaLeuAlaMetGlnSerSerAsnAlaGluAlaLeuPheG 740
||| : : : : :
7376 GAGAGCCCGCTAGAGCTGATGAAAGAAAGACAGCTC...TGTGCTTAC 7422
740 lGlnLeuAspLysLeuArg.....ThrIleTyrAspAlaTyrVal 753
||| : : : : :
7423 AAACAAGTAAGAAAGATGACATCATGACAGATTTACTTAAGGATAGATG 7472
754 LysLeuValGluGluThrIleProLeuAlaGluLysAsnLeuAsnGlnI 770
||| : : : : :
7473 GAGAACCTTGAAAGAGAGCTAGAGATAGCCAGACAAAC...CAAGAGCA 7519
770 sLeuAlaAspLysSerGlnLysAlaGlnAlaPheAspLeuLeuGly 787
||| : : : : :
7520 TGCAGCTCTTGAGGACAGAAATCCAA..... 7547
787 AlLeuAlaHisValAlGlnMetAspArgAspAlaValAlaLeuLeuGln 803
||| : : : : :
7548GGAGAGTAGAGACCTTAAAGCAAAATAGAAAGGATGACCCAA 7592
804 ProThrAspThrIleAspArgHisVal..... 812
7593 AGTGTGAGAGCTGGAATTAGATGTGTACTATAAGTCCAGAAAAAGA 7642
813HisGluIleGlnLeuValLysGluValGluAspLeuG 826
||| : : : : :
7643 AAATGTGACAAATGAAATTAACAAAAGACAGCAAGCATATCTGAAATAG 7692
826 lutyAlaLeuAspSerSerGlyArgGlyVal.....LysSerLeu 839
||| : : : : :
7693 AA...ATAATTAATTCATCATTTGAAAATTTTTCAGAAAGAAAGACAA 7739
840 GluGluIleGlnLeuGluLeuAsnPheLeuGlnArgThrArgAspThr 856
||| : : : : :
7740 GAGAAATGACAGATG.....AAAGAAAAATCAAGCAC 7771
856 uIleValGluValAspLeuArgAspGlnHisArgMetLeuAsnGlu 873
||| : : : : :
7772 TGCATGAGATG.....CTTCAACACATTAATAAGACCTTAATGAGA 7815
873 sPheSerSerAlaGlnValArgThrHisAsnAlaArgGluGlu...Lys 888
||| : : : : :
7816 GAGTGGACCCCTG.....CATATGACCAAGAGCCCTGTAAG 7853
889 ValLysLysSerSerIleLeuGluArgPheGlnLysSerGluGluLe 905
||| : : : : :
7854 GCCAAAGACAGAT..... 7868
905 uValLeuLeuAlaGluGluLysGluGlnLeuIleValGluLysLeu 922
||| : : : : :
7869CTTAGTAGTCACAGTAGAGTGTCTGAACCTTGAGAAAGCTCAGT 7911
922 euGlnGluSerLeuAspPro.....LeuSerLys 931
||| : : : : :
7912 TGCTACAGGCTTGATGAGGCCAAATAATTAATTAATGTTGTCATCT 7961
932 GluLysGluSerLeuLeuGlnGluTyrAsnAlaLeuLysGlnLysLeuAs 948
||| : : : : :
7962 TCAGTGAAGGCTCATTTCAAGAAAGTAAGAAAGGCAAGCAAGAAATGGA 8011
948 rGluGluTyrHisGlnLeuAlaGluArgLysArgGluPheGlnGlnLys 965
||| : : : : :
8012 GAAGAGGATGAAGAAATCAGTAGAGTGAATAATCAAAATTCAGACCAAG 8061
965 euAspAlaLeuGluArgLeuAsnMetLysIleLysGlyTyrLeuAspSer 981
||| : : : : :
8062 AGCAGCTTGC..... 8072
962 LysLysAsnGlnLysLysGluLeuGlnGluArgHisValLeuCysHis 998
||| : : : : :
8073TCTAACTGTGCCAGGTGGAAGAGACACACACTTTGGA 8113
998 sSerGlnLeuGlnSerCysMetalLysGlnGlnArgIleSerAlaGlu 1015

```

8114 GGAGCA.....AACTTGAACATGAAATCGCTGAAAT 8151
1015 euasnlyserlysgluLeuenglnlyglnlyglnleuLyarghpn 1031
8152 TGGAGCAGAAATGCAAGTCTAACAATCCAAA.....AATGCCCT 8192
1032 IleasparanleuLy.....TyraTgLySThrLySAI 1043
8193 TTGCGAGCAATTAAGAGTCTCAGAGTCTTCAAGAAATCTAGAGAA 8242
1043 aspyalgluInleuThraAspIlegluSerleuInluArgleuL 1060
8243 TGAGCTTAATGGACAAAATGACAAAATGCTCTTTGTTAAAAA... 8288
1060 euserIlegllySerleuSerIallegluaAAspLeuLyarghIser 1076
8289 .....GTAAACAAATGACTGCAAAAGAACTGAGCTCAGAGGAAATG 8333
1077 GlnGlu.....LySGuArgleuAsnSerGI 1085
8334 CATGAGATGGCAGCAAAACAGAGAGCTGCAGAGAACTCAGTGAGA 8383
1085 uPheasnArgTrpGlnGlyThrleuSerValTyrglnserasnlserL 1102
8384 GAAAATATAGCTAGCTGGAGAGTTCAGTCTGTGGAAAGAAATAAAG 8433
1102 ysnHleLyglnGlnleuLyserleuSerGlnTyLyAspIlegluLyarg 1118
8434 CGAGCAAAAGATCAATGTAAG..... 8453
1119 TyTrpAsnGlnPheleuGlnleuLyThrTrgIuMetAlaAsnLyas 1135
8454 .....GACCTCACACTAGAAAATAGTGA..... 8477
1135 PLeuAspArgTyTrpThrAlaLeuAspLyAlaLeuMetArgPheHLS 1152
8478 .....TTGAAAGAGAGCTTACATTCAGTCAGCACACA 8505
1152 erMetLyseMetGlnluIleAsnLyIleIleLyglnleuTrpGln 1168
8506 AAGACAGAGTGAAGAAAGAGAAAGTGAAGAAATAGCTGAATAT 8555
1169 ThrTyraGlyGlnAspIleAsp.....TyrlleSerIleAsnSerAs 1183
8556 CAGCTACGCTTCATGAAAGCTGAAAGAAACCCAGGCTTTCCTTTGA 8605
1183 pserGlnGlyAlaGlyThrArgSerTyserTyraGlyValValMetGln 1200
8606 CACA.....AACAAACAGTATGAGTAAGTAATCCAGA 8637
1200 hGlyAspAlaGluLeuGluMetArgLyArgCysSerAlaGlnLy 1216
8638 CATACCGAAGAAATGCACTCTAAGAAAGATGCTCAGTCCACAGAG 8687
1217 ValIleAlaSerleuIle 1222
8688 CTGGAAGATGACCTTTTA 8705

```

seq_name: /cgn2_6/ptodata/2/lna/5A_COMB.seq:us-08-139-937-12

seq_documentation_block:

Sequence 12, Application US/08139937

Patent No. 5821070

GENERAL INFORMATION:

APPLICANT: LEE, MEN-HWA

APPLICANT: SHAN, BEI

TITLE OF INVENTION: CELLULAR GENES ENCODING

TITLE OF INVENTION: RETINOBLASTOMA-ASSOCIATED PROTEINS

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESSES:

ADDRESSEE: CAMPBELL AND FLORES

STREET: 4370 LA JOLLA VILLAGE DRIVE

```

CITY: SAN DIEGO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/139,937
FILING DATE: 20-OCT-1993
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/979,156
FILING DATE: 20-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: CAMPBELL, CATHERYN
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-CJ 9370
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-535-9001
TELEFAX: 619-535-8949
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 4868 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-139-937-12

```

alignment_scores:

Quality	Ratio	Length	Gaps
419.50	0.655	1194	46
Percent Similarity: 53.601	Percent Identity: 20.184		

alignment_block:

US-09-538-396-2 x US-08-139-937-12

Align seg 1/1 to: US-08-139-937-12 from: 1 to: 4868

```

113 IleguSerValleuGlnThrIleAsnProHISThrGlyGluLyValcy 129
51 ATCCATATCTTCAACTCGGGTAAAGACATCAATGAAATTTAG 100
129 sleuSerTyraGlyCysAlaAspMetAspArgGluIleProAlaMetG 146
101 ATTACTTCATGTGATAGAGAGACCGGACAGAAAGTGAAGTTGCTAA 150
146 lyVal.....SerLyAlaValleuGlnAsnValIlePhe 157
151 ATGAATGAAGAAATTAAGCTCAAAATCCATTTACAGAGGTAACTA 200
158 ValHISGlnAspGluSerAsnTrpProleuGlnAsp...ProSerThrle 173
201 ATGACCAAAATTAAGATGCAATGCAATTTGAAAAAATAGTTGGGGAAT 250
173 uLyLyLyAspPheAspIlePheSerAlaThrArgTyTrpThrLySAI 190
251 TAAGAAAGAAACCTCAGATTA.....ACTGAAAAAT 282
190 euGlnValIleLyLyLeuHISLyAspGlnMetGlnluIleLyThr 206
283 TCGAATATTTTCTTGTGATCACCGAGAGTTACTCCAGAGAGTAAACT 332
207 PheArgleuLyLeuGlnAsnleuGlnThrValLyAspGlnAlaHISLy 223
333 TGTGAAGCCTAATTCGATTTAGAAATGCAATGCAGATTAATATCA.. 380
223 sleuArgIleAsnIleAlaGlnAspGlnluLySerAspAlaSer.... 238

```

381 ...CGTGAAGATATGTGAGATTAATGTGGCCAGGTGAATGACAGCTGGA 426
239LysSerGlnMetGluGlnLeuLysGluLys 248
427 AGGAGAGATTTCTTGATGTGGAATAAGTGTGATGATGATGATGATGAG 476
249 IleCysGluThrGluArgGluIleLeuGlnMetGluThrSerLeuAspGlu 265
477 AAGGTAGCATGTGAGCATGAGCCCTTACCTGAGGCTGACTTAAGGT 526
265 uLeuArgArg..... 268
527 AGTTCAAACAGAGAAGCTATGTTAGAAAAAGACAATGAAATAAGCAGA 576
269LeuGlnGluGlnIleAspIleLysAlaThrGluArg 280
577 AGTTATTTGCTGCTTGAAGAAAGCTCTCACTGCTGCAAGTGAAGA 626
281 SerThrLeu.....LeuThrGlnGlnIleGluLysLe 291
627 AACCAAGCTTCGTGAGAAATTACTACTATGTCAAAAAACCCAGGCACCT 676
291 uAlaAlaLeuSerGluGluGluAsnGluAspThrAspGluGluLeuMetGlu 308
677 GGATCAGTTGCTGTGAAAAATGAGAGAAACACAAAGAGCTGAGTCTC 726
308 rGlnIleThrLysPheGluGluArgGluIleAlaLeuGlnGluThrLysIleSer 324
727 ATCAAAGTGTGCTCTCCATTTGCTCAGGTGCGAGAGCGAGAGGTGAAG 776
325 LysLeuValArgAspMetLysAspGluAlaIleSerLysSerValLeu 341
777 GAAAGAGCGGAAGCTCTTCAAGCTTTGCTCTGATGTGATGATGATGAT 826
341 rLysGlnAsnSerGluLeuThrHisGluIleGluLysLeuGlnAlaGlu 358
827 AAAAGCAAAATCATCTCCAGAAAAAGCTGCAAGATTGTGAAAAAGACT 876
358 lAspAlaIleLeuThrMetLysHisGluArgAspSerAspIleLysAsn 374
877 CACAGCGCAGTCTTGTGACAAAAATGTGAGCTGGAACCAAAATTCACAA 926
375 IleCysThrLysHisAsnLeuGluProValProGlnHisProPheThrAs 391
927 CTGATTAAGAG.....AA 940
391 nAspValAlaMetAsnLeuThrAsnArgIleLysAlaArgLeuSerSerL 408
941 AGAATTGCTTGCAGAGATCTGAAGAGCTGCGAGGCCAGAGCTGAGTAAT 990
408 euGluAsnAspLeuLeuAspLysLysLysSerAsnGluAspGlnLeuAsp 424
991 CAGATTATGAAAGCTGAATGCTCCAAAGCCCTTGAGAGCCGCCACCTGCTG 1040
425 ValLeuThrPheLysHisThrLeuLysIleAsnAlaArgLysSerGluValAs 441
1041 GAGAAAGTGAAGTCCGATTGAGGCTGAGCTCAACACAGAGGAGAGTCA 1090
441 pGluGlnIleGlnSerLysIleGluSerMetSer..... 452
1091 T...CAGCTGAGAGAGAGCATCGAAGAACTGAGAGTTCGATGAGGCCG 1137
453GlyIleLeuArgArgArgLysAspLysGluLys 463
1138 ATGAAAGAGAGAGCTGACATCGCAGAGAACTGAAAGAACCCGAGCGG 1187
464 GluArgAspAlaAlaGluValGluLeuSerLysPheAsnLeuSerArgIle 480
1188 GAGAAAGATTCACCTTAAGTA.....AAAGT 1213
480 eAspGluArgGluArgHisMetGlnIleGluValGluArgLysThrLeu 497
1214 TGAGAAAGCTTGAAGGAGATTCAGATGTCAAGAGAAACAGAGAGCTAG 1263

497 lA...LeuGluGluArgAspLysAspSerIleIleSerGlnLysArgThr 512
1264 TGATTTCTTGATGCCGAGAAATCCAAAGCAGAAATAGAGACTTAAAGCA 1313
513 GluValLysSerLeuGlnGlnLysIleLysValLeuLeuArgGluLysAs 529
1314 CAATAGAAAGATGCGCAGAGAGCCCTGAAAGTTT.....GAATAGA 1357
529 pIleIleAsnArgAsnAlaAspGluArgValLysLeuGluLeuLysAla 546
1358 CCTGTC..... 1364
546 sPalaLeuGlnSerLysAspLysLeuAsnGluIleValAsnGluHis 562
1365 ...AGTTAAGTCTGAAAAAGAAATCTGCAAAACAAATACAGAAAAA 1412
563 LysAspLysIleLysLysValLeuArgGluArgAspProPheGluLysAs 579
1413 CAGGTGAGTGTGACAGACAGACAAAGTTACTCTCTCATTTAAAGTCT 1462
579 pMetLysLysGluIleAsnGlnAlaPheThrProValAspLysGluTyr 596
1463 GTTAGAA.....GAAAGAGAGCAG 1482
596 snGluLeuArgSerLysSerGlnGluAlaGluGlnGluLeuLysPheThr 612
1483 CAGAGATACAGATCAAA...GAAAGATCTAAACTGCGAGTGAAGTCTT 1529
613 GlnSerValThrAspAlaArgGluGlnLeuThrLysLeuArgArgAs 629
1530 CAGAAATAGTAAAGAGCAAAATGAGCGAGTGAAGCAAGCTTGTGTGTGA 1579
629 P.....MetAspAlaLysArgArgPheLeuAspSerLysLeuGlns 643
1580 CCAAGAAATATGAGGCCACAGACAGAGCTTACCCACCAATGAGG 1629
643 erIleLeuGlnIleSerAlaAsnValAspMetPheProLysValLeuGln 659
1630 AAGAGCATCAGCTGAGAAATAGCATTTGAAAGCTGAGAGCCGCCATGAA 1679
660 AspAlaMetAsnLysArgAspGluGlnLysArgLeuGlnAsnPheAlaAs 676
1680GCTGATGAAAGAGACAGCTC..... 1700
676 nGlyMetArgLysMetLeuAlaProPheGlnHisLeuAlaArgLysAsn 693
1701 TGATGCTTCAACAACTGAAGAGAAAGTGAAGT.....C 1734
693 lAsValCysProCysCysGluArgAlaPheThrProAspGluGluAspGlu 709
1735 ATCCAGATTACTTAAGGTAGAGTGAAGAACTTGAAAGAGAGCTAGAG 1784
710 PheValLysArgGlnArgMetGlnAsnSerSerThrAlaGluArgSerLys 726
1785 ATAGCCAGGACAAACCAAGAGCATGCACTTGAAGGCAAGAAATTCGCA 1834
726 sAlaLeuAlaMetGlnSerSerAsnAla.....GluAlaLeuPheGln 741
1835 AGGA...GAGGTAGAGACCTTAAGCAAAATAGAAAGGAGTAGACCAA 1881
741 lLeuAspLysLeuArgThrIleThrAspAlaArgValLysLeuValGlu 757
1882 GTCTGAGAGTCTG.....GATTTAGAGTGT 1907
758 GluThrIleProLeuAlaGluLysAsnLeuAsnGlnHisLeuAspGlu 774
1908 GTTACTTAAAGTGCAGAAAAAGAAATCTGACAAATGATTAACAAAAAGA 1957
774 uSerGlnLysAlaGlnAlaPheAspLeuLeuGluValLeuAlaHisV 791
1958 GCAAGAGCGAATATCGAATTAGAAATTAATAATTTGATTTGAAATAT 2007

```

791 aGlnMetAspArgSP..... 796
2008 TTTTCCAGAAAAAGAGCAAGAGAAATGACATGACAAAGAAATATCAAGC 2057
797 ..AlaValGluAlaLeuLeuGlnProThrAspThrIleAspArgHisVa 812
2058 ACTGCCATGAGATGCTTCAACACAAATTAAAGAGCTCAATGAGAGAGT 2107
812 LHisGluLeu.....GlnGln 818
2108 GGCACCCCTGATATGACCAAGAGCCTGTAAAGCCAAAGAGAGAGATC 2157
818 euValLysGluValGluAspLeuGlnTyrAlaLeuAspSerSerGlyArg 834
2158 TTAGTAGTCATAGTAGTGTCTTGACTT.....GAGAGGCTCTAG 2198
835 GlyValLysSerLeuGlnGlnLeuLeuLeuAsnPheLeuGlnArg 851
2199 TTGCTACAAAGCCTGTATGAGGCCAAATAATATATTGTTTTCATATC 2248
851 GThrArgAspThrLeuIleValGluValAspAspLeuArgAspGlnHisA 868
2249 TTCAGTGAATGGCTCAATCAAGAGTAGAAGTGGCAAGCAAGAACTGG 2298
868 tGMetLeuAsnGluAspMetSerSerAlaGlnValArgTrpHisAsnAla 884
2299 ACAAGAGAGATGAGAAATCACTAGCTGAAATCAATTCACAAACCA 2348
885 ArgGluGluLysValLysAlaSerSerIleLeuGlnArgPheGlnLysE 901
2349 GAGCAGAGCTGTCTTAACCTGCTCCAGGTGGAAGAGAGACCAACTTGG 2398
901 tGlnGluGlu.....LeuValLeuLeuAlaGlnGlu 912
2399 GAGAGAGCAAACTAGAACTGAGAAATCTGACAGTGAATTGGAGAGAGA 2448
912 ySGLuGlnLeuIleValGluLys..... 919
2449 ACATCCATGCTACATCCAAATGCCTCTTGCAGACATCATAGAA 2498
920 .....LysLeuLeuGlnGlnGlnSerLeuAspProLeuSe 930
2499 GTCGTGACAGATTCTTACAAAGATCTAGAGATGAGCTTGAATTCAGAA 2548
930 rLySGLuLysGlnSerLeuLeuGlnGlnTyrAsnAlaLeuLysGlnLys 946
2549 AATGGCAAAATGCTCTTGTGAAAGTAAACAAATGACTGCAGAAAGG 2598
947 .....LeuAspGlnGluTyrHisGlnLeuAlaGluArgLysArgGlu 960
2599 AACTGAGCTGCAAGAGGAATGATGATGACACAGAAACACAGCAGAG 2648
961 PheGlnGlnGlnLeuAspAla.....LeuGlyArgLeuAs 972
2649 CTGCAAGAGAACTAGTGAAGAGAAATAGCTAGCTGAGAGAGTTGCA 2698
972 nMetLysIleLysGlyTyrLeuAspSerLysAsnGlnLysLeuLysG 989
2699 GTTACTGTGTGGAAGAAATTAAGAGCAGCAA.....GATCAATGGAAG 2742
989 LLeuGlnGlnLysHisValLysSerGlnLeuGln..... 1002
2743 AGCTC.....ACACTAGAAATATGAAATTGAAGAAGAGCCTA 2780
1003 SerCyMetAlaLysGlnGlnArgIleSerAlaGlnLeuAsnLysSerLy 1019
2781 GATGTGATCACAAGAGCAG...GTGGAAGAGAGAGGAAGAGTGAAGA 2827
1019 sGluLeuLeuGlnGlnGlnGlnLeu.....LysArgAsn 1031
2828 GGAAATAGCTGATATCACTAGGCTTCATGAGCTGAAAGAAACACCC 2877
1032 .....IleAspAsnLeuLysTyrArgLysThrLysAlaAsp 1044

```

```

2878 AGCCTTCTTTTGGACACAAACACAGTAT.....GAAGTAGAA 2918
1045 ValGlnGlnLeuThrArgAspIleGluSerLeuGlnGlnArgLeuSe 1061
2919 ATCCAGATACCGAGAAATTAAGCTTCTTAAGAGATGCTC..... 2963
1061 rIleGlySerLeuSerAlaIleGluAlaAspLeuLysArgHisSerGln 1078
2964 .....AGTCACAGAAAGCTGAGATGAGCTTTTAAGTCTAGT... 3002
1078 LulysGluArgLeuAsnSerGlnPheAsnArgTyrGlnGlyThrLeuSer 1094
3003 ..AAGAGAGCTCAAT.....AATCATGCAAGCTACTACTAG 3041
1095 ValTyrGlnSerAsnIleSerLysHisGlnGlnLeuLysLeuSerG 1111
3042 ATTTTG.....GAGAAATGGAAGAAACCA 3067
1111 nTyrLysAspIleGluLysArgTyrThrAsnGlnPhe..... 1123
3068 GATGACATCTA.....AATATGTAATCAGTTGAAGAAGAAATG 3111
1124 .....LeuGlnLeuLysThr..... 1128
3112 AACGTCCCGAGGGAATGAAGTTGTGATCAATCTGTAACAGCTG 3161
1129 .....ThrGluMetAlaAsnLysAspLeuAspArgTyrTyrThrAl 1142
3162 GAAGAGAAAGAGATCTCAGAGAAAGAACTCTCACTCAAGCTGAG 3211
1142 aLeuAspLys.....AlaLeuMetArgPheHisSerMetLysM 1155
3212 ACAGAGAGAGCAAGAAACAGTACTGTATG.....GATACCAAG 3252
1155 eGluGluLysLeuAsnLysIleIleLysGlnLeu 1165
3253 TCGATGAATTAACAACCTGAGATCAAGAAAGCTG 3284

seq_name: /cgn2_6/prodata/2/lna/PCTUS_COMB.seq.PCT-US93-11310-12

seq-documentation_block:
: Sequence 12, Application PC/TUS9311310
: GENERAL INFORMATION:
: APPLICANT: BOARD OF REGENTS OF THE UNIVERSITY OF TEXAS SYSTEM
: TITLE OF INVENTION: RETINOBLASTOMA-ASSOCIATED PROTEINS
: NUMBER OF SEQUENCES: 14
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: CAMPBELL AND FLORES
: STREET: 4370 LA JOLLA VILLAGE DRIVE
: CITY: SAN DIEGO
: STATE: CALIFORNIA
: COUNTRY: USA
: ZIP: 92122
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US93/11310
: FILING DATE: 19-NOV-1993
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: CAMPBELL, CATHRYN
: REGISTRATION NUMBER: 31,815
: REFERENCE/DOCKET NUMBER: PP-CJ 9790
: TELEPHONE: 619-535-9001
: TELEFAX: 619-535-8949
: INFORMATION FOR SEQ ID NO: 12:
: SEQUENCE CHARACTERISTICS:

```

LENGTH: 4868 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 PCT-US93-11310-12

alignment_scores:
 Quality: 419.50 Length: 1194
 Ratio: 0.655 Gaps: 46
 Percent Similarity: 53.601 Percent Identity: 20.184

alignment_block:
 US-09-538-396-2 x PCT-US93-11310-12

Align seg 1/1 to: PCT-US93-11310-12 from: 1 to: 4868

113 ILeuSerValLeuGlnThrIleAsnProHisThrGlyLysValCys 129
 51 ATCCATAATCTTCAACCTCGGGTAAAGACATCAATAAGAGATTGAG 100
 129 sLeuSerTyrArgCysAlaAspMetAspArgLuiLeuProAlaLeuMetG 146
 101 ATTACTCATGTGATAGAGACCGCTGACAGAAAGTTGAAAGTTGCTTA 150
 146 IyVal.....SerLysAlaValLeuGluAsnValIlePhe 157
 151 ATGAATGAAGAAGATTAGACTCAAAACCTCCATTACAGAGAGTACAACTA 200
 158 ValHisGlnAspGluSerAsnTyrProLeuGlnAsp..ProSerThrLe 173
 201 ATGACCAAAATTGAAGCATGCATAGAAATGAAAAAATAGTTGGGGAAGT 250
 173 uLysLysLysPheAspAspIlePheSerAlaThrArgTyrThrLysAlaL 190
 251 TAAGAAAGAAACTCAGATT.....AGTGAATAAT 282
 190 euGluValIleLysLysLeuHisLysAspGlnMetGlnLuiLeuYthr 206
 283 TGGAAATATTTTCTGTGATCACAGAGACTTACTCCAGAGAGTAGAAACT 332
 207 PheArgLeuLysLeuGlnAsnLeuGlnThrValLysAspGlnAlaHisLys 223
 333 TCTGAAGCGCTCAATCTGATTAGAAATGATGCAGATAAATCAACA... 380
 223 sLeuArgGluAsnIleAlaGlnAspGlnLuiLysSerAspAlaSer.... 238
 381 ...CGTGAAGATATTGGAGATAATGTGCCCAAGTGAAATGACAGCTGGA 426
 239LysSerGlnMetGlnLuiLeuLysGluLys 248
 427 AGGAGAGATTTCTGTGATGGAATAATGAGCTGAGTAGATCAGATCGAG 476
 249 IleCysGlyThrGluArgGluIleLeuGlnMetGluThrSerLeuAspIle 265
 477 AAAGCTAGCATGAGCATGAAGCCCTCTACCTGAGCGTGCATTAGAGGT 526
 265 uLeuArgArg..... 268
 527 AGTTCAAAACAGAAAGCTATGTTAGAAAAAGACAATGAATAATACACA 576
 269LeuGlnGlyGlnIleAspIleLysAlaThrGluArg 280
 577 AGGTATTCTGCTGCTTGAAGAAGACTCTCAGTGTCTACAAAGTAGAGAG 626
 281 SerThrLeu.....LeuThrGlnGlnHisGluLysLe 291
 627 AACGAGCTTCTGAGAGATTAGACTATGTCAAAAAACCAGCGCAGCT 676
 291 uAlaIleLeuSerGluGluAsnGluAspThrAspGluLuiLeuMetGluT 308
 677 GGATCAGTTGTCTGAAAAAATGAGAGAAACACAAGAGCTTGAAGTCTC 726

308 rPcInThrLysPheGluGluArgIleAlaLeuLeuGlnThrLysIleSer 324
 727 ATCAAGTAGAGTGTCTGCATTCATTCAGGTGCGACAGCAGAGGTGAAG 776
 329 LysLeuValArgAspMetAspAspIleAlaSerTyrSerSerValLeuSe 341
 777 GAAAGACGGAACCTCTCAGACTTGTGCTCTGATGAGAGACTCTT 826
 341 rLysGlnAsnSerGluLeuThrHisGlnLuiLeuLysLeuGlnAlaGlu 358
 827 AAAAGCAAACTCATCTCCAGAAAGCTGCAAGATTGGAAAAAGACT 876
 358 LaAspAlaHisLeuThrMetLysHisGluArgAspSerAspIleLysAsn 374
 877 CACAGGCACTGTCTTGACAAAATGTGACCTGGAACCAATTCACACA 926
 375 IleCysThrLysHisAsnLeuGlyProValProGluHisProPheThrAs 391
 927 CTGAATAAAGAG.....AA 940
 391 nAspValAlaMetAsnLeuThrAsnArgLileLysAlaArgLeuSerSerL 408
 941 AGAATTGCTTGTCAAGGAATCTGAAGCCTTGAGCGCCAGACTGAGTGAAT 990
 408 euGluAsnAspLeuLeuAspLysLysLysSerAsnGluAspGlnLeuAsp 424
 991 CAGATTATGAAGCTGAATGTCTCCAGAGCCTTGAGAGGCCCATGAG 1040
 425 ValLeuThrLysHisTyrLeuLysIleAsnAlaArgTyrSerGluValAs 441
 1041 GAGAAAGTGAGATGCCATTCAGAGCGTCAACACAGAGAGAGATGCA 1090
 441 pGlyGlnIleGlnSerLysIleGluSerMetSer..... 452
 1091 T...CAGCTGAGAAAGAGCATGAGAACTGAGAGTTCCCATGAGCGCG 1137
 453GlyIleLeuArgArgArgLysAspLysGluLys 463
 1138 ATGAAAGAAAGCAGCTGCACATCGCAGAGAACTGAAAGACCGACGCG 1187
 464 GluArgAspAlaAlaGluValGluLeuSerLysPheAsnLeuSerArgIle 480
 1188 GAGAAATGATTCATTAAGTA.....AAAGT 1213
 480 eAspGluArgGluArgHisMetGlnIleGluValGluArgLysThrLeuA 497
 1214 TCGAAGACCTTGAAGGGAATGTCAGATGTCAGAAAGAAACCGAGAGCTAG 1263
 497 Ia...LeuGluArgAspTyrAspSerIleIleSerGlnLysArgThr 512
 1264 TGATTCTGATGCCGAGAAATCCAAAGCCAGAACTAGAGACTTAATAACA 1313
 513 GluValTyrSerLeuGlnLuiLysIleLysValLeuLeuArgGluLysAs 529
 1314 CAATAGAGAGATGGCCAGAAACCTGCAAAACAAATACAGAAAAA 1412
 529 pIleLeuAsnArgAsnAlaAspIleArgValLysLysLeuLysLysAs 546
 1358 CTTGTC..... 1364
 546 sPAlaLeuGluSerSerLysAspLysLeuAsnGluLuiLeuValAsnGluHis 562
 1365 ..ACGTTAAGCTGAAAAAGAACTGCAAAACAAATACAGAAAAA 1412
 563 LysAspLysIleLysLysValLeuArgGlyArgAsnProPheGluLysAs 579
 1413 CAGGTCAGTGTGCAGAACTAGCAAGTAAGTCTCTCATTTAAAGTCT 1462
 579 pMetLysLysGluIleAsnGlnAlaPheTyrProValAspLysGluTyrA 596
 1463 GTTAGAA.....GAAAGGAGACAAAG 1482

596 sngluearqserlyserglnlualaglnlueuylsapherthr 612
1483 CAGAGATACATCAAA...GAGATCTAAAGTCGAGATGCTT 1529
613 glnserlyvalthrappalaarglnlueuthrlyleuargllys 629
1530 CAGATCATGTTAAAGAGCTAAATAGAGCATGACACCTGTGTGCGA 1579
629 p.....Metaspalaalysargphleuaspserlyseu 643
1580 CCAAGAAATTTGAGGCCACAGAGCTGTAGACCCACCAATTAAGG 1629
643 erlleuaglnlleserlalsanvalaspmetpherolysvalleu 659
1630 AAGACATACATGAGAAATAGCATTTGAAAGCTGAGAGCCCGCTAGA 1679
660 Aspalamelaalysargaspglulnlyargleuuluanphealals 676
1680GCTGATGAAAGAGCAGCTC..... 1700
676 nglmetarggluethleualarprophelnhlleualarglysanh 693
1701 .TGCTCTTACACACTGAGAGAAAGTGCAT.....C 1734
693 lvalcysercyasgluargalapherthrproaspglulnlyasp 709
1735 ATCGAGATTCTTAAGGTTAGAGTGAGAACCTTGAAAGAGCTAGAG 1784
710 phevallylsglnargmetglnabnserserthrallagluargserly 726
1785 ATAGCCAGACAAACCAAGAGCATGACCTGTGAGCAGAGAAATTCGA 1834
726 salaleualametgluserasana.....Glualeuphegln 741
1835 AGGA...GAGGTAGAGCCCTAAAGCAAAATAGAGGATGACCCAA 1881
741 luleuasplysleuargthrillerthasplalyvallyleuvalglu 757
1882 GTCTGAGAGGCTG.....GAATTAGATCTT 1907
758 gluthrileproleualaglnlyasleuabnglnhlleualaspl 774
1908 GTTACTATTAAGGTCAAGAAAGAAATCTGCAATATTAACAAAGA 1957
774 userglnlysalaglnalaphenaspheleucllyalaleualahsv 791
1958 GCAGAGCGCAATATCTGAATTAAGAAATTAATCATCATTTGAATA 2007
791 alglmetasplargsp..... 796
2008 TTTGCAAGAAAGAGCAAGAGAAATGACAGTAAAGAAATCAAGC 2057
797 ...AlaValgluaaleuclnleuclnprothrapsrthrleasparghsya 812
2058 ACTGCCATGAGATGCTCAACACAAATTAAGAGCTCATGAGAGCT 2107
812 lhlleuclle.....glncln 818
2108 GGCACCCCTGCAATATGACCAAGAGCCTGTAGGCCAAGAGCGAATC 2157
818 euvallygluvalgluaspheugluytralaaleuaspsersegllyarg 834
2158 TTAGTATGATGAGTGTCTGAACTT.....GAGAGGCTGAG 2198
835 glyvallyserseuclnleuclnleuclnleuclnleuclnleucln 851
2199 TTGCAACAGAGCTGATGAGGCCCAAAATATATATGTTTTCATATC 2248
851 gThrargaspthrleuclleuclnleuclnleuclnleuclnleucln 868
2249 TTCAGTATGAGCTCATTAAGAGTGAAGATGCGCAAGAGAAATGCG 2298
868 rgmetleuabnglnuaspmetseralaglnvalargtrphlsanla 884

2299 AGAAGAGATGAGAAATCATGACTGAAATCAATTCAGAGCA 2348
885 arglglulnlyvallysalaserleuclnleuclnleuclnleucln 901
2349 GAGAGCTGTCTCTAAATGCTGCCAGGTGAGAGAGAGCAGCACTT 2398
901 rglulnleuclnleuclnleuclnleuclnleuclnleuclnleucln 912
2398 GAGAGCAAACTTAGACTGAGAAATCTGACAGTGAATGAGAGCA 2448
912 ysglnleuclnleuclnleuclnleuclnleuclnleuclnleucln 919
2449 AGATCAAGTGTCTCAATCCAAATGCTCTTTCAGAGACATTAAGAA 2498
920Lysleuclnleuclnleuclnleuclnleuclnleuclnleucln 930
2499 GTGCTGAGACTTCTTACAAATCTAGAGAAATGCTGATTAAGCAAA 2548
930 rlysglnlysgluserleuclnleuclnleuclnleuclnleucln 946
2549 AATGACAAATGCTCTTGTGAAAAAGTAAACAAATGACTGCAAGG 2598
947Leuaspglulnlythrlysglnleualagluarglysarg 960
2599 AAATGAGCTCAGAGGAGAAATGATGATGATGATGATGATGATGAT 2648
961 pheglnleuclnleuclnleuclnleuclnleuclnleuclnleucln 972
2649 CTGCAAGAGAACTGAGGAGAGAAATGAGCTGAGCTGAGAGTGA 2698
972 nmetlyllysglytrleuaspserlyllysanglnlyseuclnlysg 989
2698 GTTACTGTGAGAAATTAAGAGCAGAA...GATCAATTGAGAG 2742
989 luleuclnlyarghllysalaleucylserleuclnleuclnleucln 1002
2743 AGCTC.....ACACTGAAATAGTGAATTAAGAAAGAGCCTA 2780
1003 sercysmetalalyglnleuclnleuclnleuclnleuclnleucln 1019
2781 GATTGACAGCAAGAGAGAGAG...GTGAAAGAGAGAGAGAGAGAG 2827
1019 sgluleuclnleuclnleuclnleuclnleuclnleuclnleucln 1031
2828 GGAATAGCTGAATATGCTGACGCTCATGAGCTGAGAGAGAGAGAG 2877
1032Ileaspasplnleuclnlyarglythrlysalaspl 1044
2878 AGGCTTGTCTTGTGACCAAAAGAGAGTAT.....GAAGTACAA 2918
1045 valglulnleuclnleuclnleuclnleuclnleuclnleuclnleucln 1061
2919 ATCCAGACATACCGAGAGAAATGCTTCAAGAGAGAGTGTCTC 2963
1061 rllleuclnleuclnleuclnleuclnleuclnleuclnleuclnleucln 1078
2964AGTTCACAGAGAGTGTGAGAGTGTCTTAAAGCTTACT 3002
1078 lulysglnlyargleuclnleuclnleuclnleuclnleuclnleucln 1094
3003 ...AAAGAGAGCTCAAT.....AATCATGAAAGCTACTACTGAG 3041
1095 vallyrglnserasnleuclnleuclnleuclnleuclnleuclnleucln 1111
3042 ATTTTG.....GAGCAATTAAGAGAGAGAGAGAGAGAGAGAG 3067
1111 ntyllyaspillegluyargtrthrlysalnleuclnleuclnleucln 1123
3068 GATGACATATCA...AATATGTAATCACTGAGAGAGAGAGAGAGAG 3111
1124leuclnleuclnleuclnleuclnleuclnleuclnleucln 1128

3112 AACGTGCCAGGAGAAATGAAGTGTGTGATCAATCTGTAACAGCTG 3161
1129ThglumetalasnllyspaleusparqTYrTYrThAl 1142
1162 GAAGAGCAAAAGAGATGACTGACAGAAAGACTCTCTCAACTTCAAGCTGC 3211
1142 aleuasplys.....AlaleumetcarphhehlssemeltySM 1155
3212 ACAGAGAGAGCAAGAAACAGCTACTGTTATG.....GATACCAAGG 3252
1155 etglugluileasnllysllellelysgluLeu 1165
3253 TCGATGATTTAACAACGTGATCAAGAGACTG 3284

seq_name: /cgn2_6/ptodata/2/lna/5B_COMB.seq:US-08-687-080-54

seq_documentation_block:

Sequence 54, Application US/08687080
Patent No. 5965427

GENERAL INFORMATION:

APPLICANT: Gregory Dolganov

TITLE OF INVENTION: Human RAD50 Gene and Methods of Use Thereof

NUMBER OF SEQUENCES: 175

CORRESPONDENCE ADDRESS:

ADDRESSEE: Dehlinger & Associates

STREET: 350 Cambridge Avenue, Suite 250

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94306

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/687,080

FILING DATE: 17-JUL-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/592,126

FILING DATE: 26-JAN-1996

ATTORNEY/AGENT INFORMATION:

NAME: Sholtz, Charles K.

REGISTRATION NUMBER: 38,615

REFERENCE/DOCKET NUMBER: 4600-0111.30

TELEPHONE: (415) 324-0880

TELEFAX: (415) 324-0960

INFORMATION FOR SEQ ID NO: 54:

SEQUENCE CHARACTERISTICS:

LENGTH: 543 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cDNA

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

INDIVIDUAL ISOLATE: RAD50-3

US-08-687-080-54

alignment_scores:

Quality: 391.50

Ratio: 2.663

Percent Similarity: 79.032

Percent Identity: 43.548

alignment_block:

US-09-538-396-2 x US-08-687-080-54

Align seg 1/1 to: US-08-687-080-54 from: 1 to: 543

71 VALAAGLYglutThrlysglygluilelyleuArpPheLySth 87
1 GTTGTCAAGAAACAGATGTGAGACCCAGATTGCTGCAATTCCGCGA 50
87 TALAAGLYlyAspValValCysIleArgSerPheGlnLeuThrGlnL 104
51 TGTCAATGAGCAACTTATAGTGTCAAAAGATCTATGCTGTCTACTGCA 100
104 ysalSerlysmetGluPheLySalalIeGluSerValLeuGlnThrIle 120
101 AAAGCAAAAGACAGAAATTTAAACTGTGGAAGAGTCAAT...ACTGCA 147
121 AasnProHstThGlygluLyValCysLeuSerTYrArgCysAlaAspMe 137
148 ACAAAAGCAT...GTGCAAAAGCTCAGTCTGAGCTCTAGTGTGAGAAAT 194
137 TAsparGlyIleProAlaLeuMetGlyValSerlySalValIleGluAl 154
195 TGACCGAGAAATGATCAGTCTCTGGGGTTCCAAAGCTGTGCTAAATA 244
154 srValIlePheValHisGlnAspGluSerAsnTrpProLeuGlnAspPro 170
245 ATGTCATTTTCTGTCATCAAGAAAGATTCTTAATTGGCTTTAAGTAAGCA 294
171 SerThrLeuLySalysPheAspIlePheSerAlaThrArgTYrTh 187
295 AAGGCTTGAGCAAAAGATTGATGAGATTTTTCAGCAACAAGTACTACT 344
187 rLySalaleuGluValIleLySalysLeuHisLyAspGlnMetGlnIui 204
345 TAAAGCTTAGAAACACTCGGAGAGTACTGAGACACAAAGTCAAGAAAG 394
204 IelYThrPheArgLeuLyLeuGluAsnLeuGlnThrValLyAspGln 220
395 TAAAGAAATATCAATGAGACTAAATATCTCAACCAATATTAAGAAATA 444
221 AlaHisLyLeuArgGluAsnIleAlaGlnAspGlnIuLySerAspAl 237
445 GCTTGAGATTCGTGATCAGATTACA.....AG 473
237 aserLySerGlnMetGluGlnLeuLyGluLyIleCysGlyThrGluA 254
474 TAAGAGCCCAAGTTACATCTTCAAGAAATTTGCAATCTTATGANA 523
254 rGlyIuile 256
524 ATGAACTT 531

seq_name: /cgn2_6/ptodata/2/lna/5A_COMB.seq:US-08-466-390-3

seq_documentation_block:

Sequence 3, Application US/08466390

Patent No. 5686562

GENERAL INFORMATION:

APPLICANT: TOKRATLY, GARY

APPLICANT: LIDGARD, GRAHAM P

TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE

INTERIOR NUCLEAR MATRIX

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: TESTA, HURWITZ & THIBEAULT

STREET: 125 HIGH STREET

CITY: BOSTON

STATE: MA

COUNTRY: USA

ZIP: 02110

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/466,390

FILING DATE: 06-JUN-1995
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: PITCHER ESO, EDWARD R.
 REGISTRATION NUMBER: 27,829
 REFERENCE/DOCKET NUMBER: MTP-013
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 248-7000
 TELEFAX: (617) 248-7100
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 6306 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1..6306
 PUBLICATION INFORMATION:
 AUTHORS: COMPTON, DUANE A
 AUTHORS: SZILAK, ILYA
 AUTHORS: CLEVELAND, DON W
 TITLE: PRIMARY STRUCTURE OF NUMA, AN INTRANUCLEAR
 TITLE: PROTEIN THAT DEFINES A NOVEL PATHWAY FOR
 TITLE: SEGREGATION OF PROTEINS AT MITOSIS
 JOURNAL: J. Cell Biol.
 VOLUME: 116
 PAGES: 1395-1408
 DATE: 1992
 US-08-466-390-3

Alignment scores:
 Quality: 376.50 Length: 1486
 Ratio: 0.510 Gaps: 62
 Percent Similarity: 49.664 Percent Identity: 19.314

Alignment block:
US-09-538-396-2 x US-08-466-390-3

Align seg 1/1 to: US-08-466-390-3 from: 1 to: 6306

103 GlnLysAlaSerLysMetGluPheLysAlaIleGluSerValLeuGln... 118
 802 CAGGGGCCAGCCACTGTGAGCCAGAGAGCTGTGAGAGCTGCGTACAA 851
 119ThiLeaSPROHisThrGlyGluLysValCys... 129
 852 GAATAGAGCCCTTACCATGCGGCTGCATGAACCTGAAAGCAGTGCACAG 901
 130 ..LeuSerTyrArgCysAlaAspMetAspArgGluIleProAlaLeuMet 145
 902 ACCTGAGAGAGAGAGAGAGAGAGATGATCCAAATC..... 939
 146 GlyValSerLysAlaValLeuGluAsnValIlePheValHisGlnAspG 162
 940AACCAAGCTTCGAGGAGGAGATGAGAGA 965
 162 userAsnTrpProLeuGlnAspProSerThrLeuLysLysPheAspA 179
 966 CTTTCCTTTAGCTGCGGAGATTGCCAGTCATCTGCAGCAGCTACAGG 1015
 179 spLlePheSer.....AlaThrArgTyrThrLysAla..... 189
 1016 ATGCCCTCAATGAGCTGAGCGAGGAGACAGCAAGCCACTCAGAGAGTGG 1065
 190 LeuGluValIleLysLysLeuHisLysAsp.....GlnMetGlnG 203
 1066 CTAG 1115
 203 uLleLysThrPheArgLeuLysLeuGluAsnLeuGlnThrValLysAspG 220

1116 CAGAAATGCTTGAAGAGAAAGAAATCTTCAG.....GGA 1156
 220 InAlaHisLysLeuArgGluAsnIleAlaGlnAspGlnGluLysSerAsp 236
 1157 AACTTCACAGCTGAGAGAACACTTGTCCAGCTGCAGAGATACACCC 1206
 237 AlaSerLysSer.....GlnMetGlnGlnLeuLys 246
 1207 CAG 1256
 246 \$GluLysIleCysGlyThrArgGluArgGluIleLeuGlnMetGluThrSerL 263
 1257 GCAAG 1306
 263 euAspGluLeuArgArgLeuGlnGlnGlnIleAspIleLys...AlaThr 278
 1307 TAGAGATGCTGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 1356
 279 GluArg.....SerThrLeuLeuThr 285
 1357 GAGCGGGGCCACTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1406
 285 rGlnGlnHisGluLysLeuAlaAlaLeuSerGluGlnAsnGlnAspThrA 302
 1407 TGACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1456
 302 spGluGluLeuMetGluTrpGlnThrLysPheGluGluArgIleAlaLeu 318
 1457 AGCAGGCTCCAGGCTCATGGGGCCGGTGTGACGTGCCAGGTGGCTCT 1506
 319 LeuGluThrLysIleSerLysLysLeuValAlaArgAspMetAspGluAla 335
 1507 CTGACCTGTGAGCTACACCACTCAATCCACCACTCAGAGAGAGAGATCA 1556
 335 rTyrSerSerValLeuSerLysGlnAsnSerGluLeuThrHisGluIleG 352
 1557 AGAAGCTGCTGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1606
 352 LysLys...LeuGlnAlaGluAlaAspAlaHisLeuThrMetLysHis... 366
 1607 CACAGAGCTCCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1656
 367GluArgAspSerAspIleLysAs 374
 1657 GTGAG 1706
 374 nLecTyrThrLysHisAsnLeuGlyProValProGluHisProPheThrA 391
 1707 GTAGCGGAGAGAG.....CAGAGAGCACTAGAGC 1735
 391 snAspValAlaMetAsnLeuThrAsnArgIleLysAlaArgLeuSerSer 407
 1736 AGGAG 1785
 408 LeuGluAsn.....AspLeuLeuAspLysLys 416
 1786 TTAAG 1835
 416 sLysSerAsnGluAspGlnLeuAspValLeuTrpLysHisTyrLeuLysI 433
 1836 GAAGGCT.....GCCAAGCTGAGAGATTCGAGAGAGAGAGAGAGAG 1879
 433 LeuAsn.....AlaArgTyrSerGluValAspGlyLys... 443
 1880 CTAATGAAG 1929
 444IleGlnSerLysIleGlnSerMetSerGlyIleLeuArg 456
 1930 GAG 1979
 456 g.....ArgArgLysAspLysGluLysGluArgAspAlaAlaGluValG 471
 1980 GACAGCCGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2029

704 roAspIuSgLuAspIuSphEVallySlySgInAryg...MetGInAsnSer 719
 2834 GAGACAGACAGCCCGGAGGCTGGAGAAAGCAACAGGAGCCCACTTCTGC 2883
 720 SerThrAlaGluArgSerLysAlaLeuAlaMetGInSerSerAsnAlaG1 736
 2884 AGCAACACAGGCGCCCTGCAGGCTTTGAGAGGGAG...GCAGA 2924
 736 uAlaLeuPheGInGInLeuAspLysLeuArgThrIleTyrAspAlaTyrV 753
 2925 GCAGATGGGCATATGATGCTGCAACGCGCTCGG...G 2956
 753 aLlySLeuValGInGluThrIleProLeuAlaGluLysAsnLeuAsnGln 769
 2957 CCGCCTGTGTGAGAGCCAGGGCGCAGCAGAGAGAGACGtGGCGCAGG 3006
 770HisLeuAlaAspGInSerGInLysAlaGlnAla.. 780
 3007 GAAGAGAGAGTGGCGCGGCTGACCCAGAGAGCGGGCGCTGCCCAAGCTGA 3056
 780 780
 3057 CGTGGCCCTGGAGAAAGCGGCCAGAGCAGAGCTTGAGATGGCGTGCAGA 3106
 781PheAspAspLeuGluGlyValLeu 788
 3107 AGCCCTCAACAGAGCAGCGCTGTGAGTTGCTGCATCCCAACAGAGCACTG 3156
 789 AlaHisValGlnMetAspArgAspAlaValAlaGluAlaLeuGlnProTh 805
 3157 GCTCATATGCTCTGACGGAAAGAGAGGCAAGAGCCAGAGTTGGCCAACT 3206
 805 rAspThrIleAsp.....ArgHisValHisGluIleGlnGlnL 818
 3207 TGTGTGTCTGGAGGAGCCCGACATTAAGAAGATGTGAGAGAACTTCGGCAA 3256
 818 euValLysGluVal.....GlnAspLeuGlnTyrAla 828
 3257 CCGTGTAGACAACTGTGAAGAAACAGCTGCTAGAAAAAGAAAGAGACGCGCA 3306
 829 LeuAspSerSerGlyValGlyValLysSer..... 838
 3307TCTGTGCTCAGAGGCCCAATGTGAGCTGCTGGCGAGACAGA 3344
 839LeuGluGlnIleGlnLeuGlnLeuAsnPheLeuG 850
 3348 GCCAAGAGGCCCAACAGCTGGAAGCACTGCGGGCAGAGTGAGACAGACTGG 3397
 850 lAsArg..... 851
 3398 AACAGCAATGCCAGAAAGCAGAGACAGAGCAGCTGACAGCCTGGAAACGAC 3447
 852ThrArgAspThrLeuIleValG1 859
 3448 CTCGAGAGCTGAGCGGCCTCCGCGGCTGAGCGGAGACAGTGTCTGTGAGAC 3497
 859 uValAsp...AspLeuArgAspGlnHisArgPheLeuAsnGluAspMetS 875
 3498 TCTGCAGGCGCAATTTGAGAGAGAGGCCAGAGAGCTAGGGCAGACAGACAGA 3544
 875 er.....SerAlaGln.....Val 879
 3548 GTGCCTTACGCTCTCGGCCCAACGGAGACTGAGCTGCTCCGACCAAGAGTA 3597
 880 ArgThrHisAsnAlaArgGInGluLysValLysAlaSerSerIleLeuG1 896
 3598 CAAGACCCACGCAAGGCTGAGATGTGAGTGGAAGGCCCAAGGTGGCCCGGG 3644
 896 uArgPheGlnLysSerGInGluGlnLeuValLeuLeuAlaGlnGluLysG 913
 3648 CCGGCAAGAGCGTGAAGGAAAAATAGCTCAACACAGCACTTGGAGAGAGAG 3697
 913 lncGlnIleuIleValGluLysLysLeuLeuGlnGlnIleuSerIleuAspProLeu 929

```

3698 AGGTGTCATCTGATATGCCAGGTCTCTGAGAGAGAGGGGAG..... 3741
|||||
930 SerLysGluLysGluSerLeuLeuGlnGlnLysAlaLeuLysGlnLys 946
|||||
3742 AGCAAGAGGTGAAGCGGCTGGTATGGCCGAGTCAAGAGAGAGCAAGAA 3791
946 sLeuAspGlu..... 949
|||||
3792 GCTGGAGAGAGTGCCTGCTGTCAGAGCAGAGACAGCCAGCAAGTGC 3841
949 ..... 949
3842 CAGAGCTCAGACAGCAGCTGCTGTCGGAGAGAGTGCAGAGCTCC 3891
950 .....GluTyrHisGlnLeuAlaGluArg.....LysArgG1 960
|||||
3892 GGAAGGAGAGCTGAGAAACAGCGGCTGCTCAGAGAACTGCGGAGAGA 3941
960 uPheGlnGlnLeuLeuAspAlaLeuGlyArgLeuAsnMetLysLysG 977
|||||
3942 GCTGACCTCAGAGCTGAGCTGCGAGAGAGCTGGCCAGAAATTGAAG 3991
977 LysTyrLeuAspSer.....LysLysAsnGluLysLeuLysGln 991
|||||
3992 CGTGCGAGAGAGTCTTCCAGAAAGAGAGCCCTCTCCACCTGCGAG 4041
992 GluArgHisVal..... 995
|||||
4042 CTCGAGCAGACACAGACAGCCCTGCTGAGTGTGCTGCCAGCTAA 4091
996 ....LeuCysHisSerGlnLeuGlnSerCysMetAlaLysGlnGlnArg 1011
|||||
4092 GCACCTCTGCCAGCAGCTGAGCGCCAGAGCGCTGCCAGAAAGCC 4141
1011 LeuSerAlaGlnLeuAsnLysSerLys..... 1019
|||||
4142 ACCGTGAGAGAGCTGAGAGCAGAGCAGCCCGCTGGGGAGCTCGGGCA 4191
1020 GluLeuLeuGlnGlnGlnGlnLeuLysArgHisAlaAspAsnLe 1036
|||||
4192 GAGCTGCTGCGGGCCAGCGGAGCTGGAGAGCTGATCTCGTGGCGCA 4241
1036 uLysTyrArgLysThrLysAlaAspValGlnLeuThrArgAspIleG 1053
|||||
4242 GAAGCTGCGAGAGCAGAGCAGAACAGCTCAGAGCTGGCGGAGAGAG 4291
1053 LysLeuGlnGlnLysArgLeuSerLysLeuSerAlaIleGlu 1069
|||||
4292 CCAGCTATGACAGCAGCTGAGACTGCTG..... 4320
1070 AlaAspLeuLysArgHisSerGlnGlnLysGluArgLeuAsnSerGlu 1086
|||||
4321 .....AGAGAGCCGATGCC.....CTGCGGGCAGAGGA 4349
1086 eAsnArgTyrGlnGlnLysLeuSerValTyrGlnSerAsnLysSerLys 1103
|||||
4350 GAACCGGGGCTGGCTGAG.....CGGGCCAACTTGGCCCGC 4387
1103 LysLysGlnGlnLeuLysLeuSerGlnTyrLysAspIleGlnLysArg 1119
|||||
4388 ACTTCTCGAAGTGGAGTGGACAGCCCGGAA.....AAGTAT 4428
1120 ThrAsnGlnPheLeuGlnLeuLysThr.....ThrGluMetAlaAs 1133
|||||
4429 GTCCAAAGATTGGCAGCGTACGTCATGCTGAGACCCGTCGCTGA 4478
1133 nLysAspLeuAspArgTyrTyrThrAlaLeuAspLysAlaLeuMet... 1148
|||||
4479 GGTGCAAGGAGAGACAGAGACTGCCCCGAGCTGAGAGGATGACTG 4528
1149 ArgPheHisSerMetLys.....MetGlnGluLysAsnLysIle 1161
|||||

```

```

4529 CCAAGTATGAGAGGTGCCAAGGTCAAGGTCTCTGAGAGAGCAGCGGTC 4578
1162 IleLysGlnLeuTyrGlnGlnThrTyrArgGlyGlnAspIleAspTyrI 1178
|||||
4579 CAGAGAGAGAGCAGAAACTCAGCTCCAGGTGGAGAACTGAGTAAGAA 4628
1178 eSerIleAsnSerAspSerGlnGlyAlaGlyThrArgSerTyrTyr 1195
|||||
4629 ACTGGCTGACTCTGACCAAGCCAGCAGAGTGCAGACAGAGCTGAAG 4678
1195 IyValIleMetGlnThrGlyAspAlaGluLeuGlnMetArgGlyArgCys 1211
|||||
4679 CTGTCCAGGCTCAGAGAGCGGAGAGCCAGAGAGCCAGCGCTTCAG 4728
1212 SerAlaGlnLysValLeuAlaSerLeuIleIleArgLeuAlaLeuAl 1228
|||||
4729 GCCAGCTGAATGAACTGCAAGCCAGTGTGAGCCAGAGAGAGCAGCAG 4778
1228 aGluThrPheCysLeuAsnLysGlyIleLeuAlaLeuAspGluProThr 1245
|||||
4779 TGACCACTATTAAGCTGCAG.....ATGAGAAAGCCAAA 4813
1245 hrAsnLeuAspGlyProAsnAlaGluSer.....LeuAlaAlaLeu 1259
|||||
4814 CACATTTATGATGCCAGAGAGCAGCAACCAAGAGCTGCAGAGCAGCTG 4863
1260 LeuArgIleMetGluAlaArgLysGlyGlnGlnLysPheGlnLeuIle 1276
|||||
4864 CGAGCCTGAGAGCAGCTGCAGAG.....GAAACAAAGAGCTG.... 4902
1276 IleThrHisAspGluArgPheAlaHisLeuIleGlyGlnArgGlnLeu 1293
|||||
4903 CGAGCTGAACTGTAAGCGCTGGCCATGACTACAGCAGGCTGGCTGA 4951
1293 IagLys 1295
|||||
4952 AGACCAAG 4959
seq.name: /cgn2_6/prodata/2/lna/5A_COMB.seq:US-08-470-950-3
seq.document:US-08-470-950-3
seq.sequence:US-08-470-950-3
seq.patent:US-08-470-950-3
seq.general:US-08-470-950-3
seq.applicant:US-08-470-950-3
seq.title:US-08-470-950-3
seq.number:US-08-470-950-3
seq.correspondence:US-08-470-950-3
seq.address:US-08-470-950-3
seq.street:US-08-470-950-3
seq.city:US-08-470-950-3
seq.state:US-08-470-950-3
seq.country:US-08-470-950-3
seq.zip:US-08-470-950-3
seq.computer:US-08-470-950-3
seq.medium:US-08-470-950-3
seq.operating:US-08-470-950-3
seq.current:US-08-470-950-3
seq.filing:US-08-470-950-3
seq.classification:US-08-470-950-3
seq.attorney:US-08-470-950-3
seq.registration:US-08-470-950-3
seq.reference:US-08-470-950-3
seq.telecommunication:US-08-470-950-3
seq.telephone:US-08-470-950-3
seq.telex:US-08-470-950-3
seq.information:US-08-470-950-3

```

SEQUENCE CHARACTERISTICS:

LENGTH: 6306 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 1..6306

PUBLICATION INFORMATION:

AUTHORS: COMPTON, DUANE A

AUTHORS: SZILAK, ILVA

AUTHORS: CLEVELAND, DON W

TITLE: PRIMARY STRUCTURE OF NUMA, AN INTRANUCLEAR

TITLE: PROTEIN THAT DEFINES A NOVEL PATHWAY FOR

JOURNAL: J. Cell Biol.

VOLUME: 116

PAGES: 1395-1408

DATE: 1992

US-08-470-950-3

alignment_scores:

Quality: 376.50

Length: 1486

Ratio: 0.510

Gaps: 62

Percent Similarity: 49.664

Percent Identity: 19.314

alignment_block:

US-09-538-396-2 x US-08-470-950-3

Align seg 1/1 to: US-08-470-950-3 from: 1 to: 6306

```
103 glnlysalaserlysmetgluphelayalalegluservalleu... 118
119 .....Thrlleasnprohishthrglyulysvalcys... 129
852 GATGAGACCTTACCATTGCGGTGCATGAMACCTGAAAGCAGTGCAGG 901
130 ..leuserlyrargcysalaspmetcasparglulieproalaleumet 145
902 ACCCTGAAGACAGAAAGACCCAGATGATCGCAAAATC..... 939
146 GlyValSerlysalavalleu... 162
940 .....AACCAAGCTTCGAGAGAGAAATGAGAGA 965
162 userasntpproleuglnaspproserthrlleuylslyspheaspa 179
966 CTTTCCTTAAGCTCGGAGATTGCCAGTCACTGCAGCAGCTACAGG 1015
179 splleheser.....AlathrArgTyrThrlysal... 189
1016 ATGCCCTCAATGAGCTGACGAGAGAGCAAGCAAGCCACTCAGAGAGTGC 1065
190 leuglivalallelyslslyleuHlslyasp.....Glnmetglnl 203
1066 CTAGAAAGACAGGCCAGCTGAGAAAGAGAGCTCAGCCAGCCCTGCGAGA 1115
203 uilelysthrphargleuylsleu... 220
1116 CAGAAATAGCTTGAAGAGAAAGCAAAATCTTCAG.....GGA 1156
220 lmlahlslyslsleuarglunasnleleaglinspglnlulysersasp 236
1157 AACTTTCACAGCTGAGAAACACTGTCTCCAGCTCGAGATACCCACC 1206
237 Alaserlysser.....Glnmetglnleuyl 246
1207 CAGAGAGAGGCGAGTGTCTGGGTGATCTGTGAGCTGAAACCTTGAA 1256
```

```
246 sglulysllecygltlythrgluarglulleu... 263
1257 GCAAGAGGACGACCTCTGTGCAAAACACACAGCTCCAGCAAGG 1306
263 euaspeluleuargargleu... 278
1307 TAGAGATGCTGAGACTGAGCAGGAGCCAGAGAGCCAGAGCTGTGCT 1356
279 gluarg.....Serthrlleu... 285
1357 GAGCGGGCCACTTGCAGAAAGAAAGCAGCAGCTGTCTAGCTGATCAC 1406
285 rglnglnhsglyluleu... 302
1407 TGACCTGAGAGCTCATCTCCACCTCAGCAGGCAAGAGAGCTG 1456
302 spgluleu... 318
1457 AGCAGGCTCCAGGCTCATGGGGCCGTTGACTGCCAGGCTGCTCT 1506
319 leugluthrllyserlyseuvalarguaspmetasparglunase 335
1507 CTGACCTGAGCTCAGCAGCTCAATGCCACCTCAGCAGAGATCA 1556
335 rtyrservalleuserlysglnasbergulieuthrlnsgluleg 352
1557 AGAACGGCTGGCTGAGAGAGGCAAGAGAGAGAGAGAGAGAGAGAG 1606
352 lyls...leuglnalaglu... 366
1607 CACAGACCTCCMAAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1656
367 .....GluArgAspSerAsp... 374
1657 GTGAGAGCTAAGCAGTACCTGAAAGCAGAGAGAGAGAGAGAGAGAG 1706
374 nilecsthrlysnhlsleu... 391
1707 GGTAGCGGAGAG.....CAGAGAGCACTAGGC 1735
391 snaspyalalame... 407
1736 AGGACATGCGCCAGCAAGCTGCGCAGAGAGAGAGAGAGAGAGAG 1785
408 leuglnasn.....Asp... 416
1786 TTAAGGAGAGGAGATGCGCTCAAGCAGCTGAGAGAGAGAGAGAG 1835
416 slysserasn... 433
1836 GAGAGCT.....GCCAAGCTGAGATTCGACAGCAGCACTCAGAGTGG 1879
433 leasn.....AlargTyrSer... 443
1880 CTATGAAGCCCGGAGACAGTCCAGCTCAGTCAACAGAGAGAGAGAG 1929
444 .....ileglnserlylleu... 456
1930 GAGAGAGCAGAGCTGAGCCGAGAGAGAGAGAGAGAGAGAGAGAGAG 1979
456 g.....Arg... 471
1980 GACAGCCCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2029
471 luleuser...Lys... 486
2030 AGTTGACAGCTGCGGTGAGCAGAGAGAGAGAGAGAGAGAGAGAGAG 2079
487 metc... 503
2080 GCCCAG.....GAGAGAGAG... 2094
503 rasperl... 520
```


949 949
3842 CAGAGCTGCAGAACGACGCTGCTGCGGGAGGAGGTGCAGAGCTCC 3891
950 GluTyrHisGlnLeuAlaGluArg.....LysArgG1 960
3892 GGGAGGAGGCTGAGAAACAGCGGCTTCACAGAAACCTGCGGAGCA 3941
960 upheGlnGlnLeuAspAlaLeuGlyArgLeuAsnMetLysIleLysG 977
3942 GCTGACTCAGAGGCTGAGCTGCGGAGGAGGCTGCGCAAGATTGAAG 3991
977 LysTyrLeuAspSer.....LysLysAsnGlnLysLysGlnLeuGln 991
3992 CGTGCGAGGAGAAAGTTCTTCAGAAAGAGCGGCTTCACACCTGAG 4041
992 GlyArgHisVal..... 995
4042 CTCGAGCACACGACACAGCCCTGCTGAGTGAAGTCTGCCACGTTAA 4091
996LeuCyHisSerGlnLeuGlnSerCysMetAlaLysGlnGlnArgI 1011
4092 GCACCTCTGCCAGAGCTGCAGGCGGAGCGGCTGCCAGAAAGGCC 4141
1011 LeSerAlaGlnLeuAsnLysSerLys..... 1019
4142 ACCGTGAGGAGCTGAGCAGAGCAAGCGGCTGCGGAGCTGCGGCA 4191
1020 GlnLeuLeuGlnGlnGlnGlnGlnLysArgAsnIleAspAspAla 1036
4192 GAGCTGCTGCGGCGGCGGAGGAGTTGGGGAGCTGATTCCTCGCGCA 4241
1036 ulsTyrArgLysThrLysAlaAspValGlnGlnLeuThrArgAspIleG 1053
4242 GAAGGTGCAGAGCGAGCGAAGCTGCAGAGCTGCGGCGAGAAAG 4291
1053 LysLeuGlnGlnLysLeuSerGlnTyrLysAspIleGlnLysArgTyr 1069
4292 CCAGCTATGCAGAGCAGCTGAGCATGCTG..... 4320
1070 AlaAspLeuLysArgHisSerGlnLysGlnLysArgLeuAsnSerGln 1086
4321AAGAAAGCGCATGCG.....CTGCTGCGCAGAGA 4349
1086 AsnArgTyrGlnGlnGlnLysLeuSerValTyrGlnSerAsnIleSerLysH 1103
4350 GAACCGGGGCGGCGGTGAG.....CGGGCAACCTTGGCGGCG 4387
1103 LsLysGlnGlnLeuLysLeuSerGlnTyrLysAspIleGlnLysArgTyr 1119
4388 AGTTCTGGAAGTGGATTGGACGAGCGCGGAA.....AAGTAT 4428
1120 ThrAsnGlnPheLeuGlnLeuLysThr.....ThrGlnMetAlaAs 1133
4429 GTCCAGAGGTTGCGAGCGCTAGCTGATGCTGAGACCGCTGCTGA 4478
1133 nLysAspLeuAspArgTyrTyrThrAlaLeuAspLysAlaLeuMet.... 1148
4479 GGTGCAGCGCAGAGCAGACAGACAGCTGCCGGGAGCGAGGATGACG 4528
1149 ..ArgPheHisSerMetLys.....MetGlnGlnLeuAsnLysIle 1161
4529 CCAAGTATGAGGTGCCAAGGTCAAGCTCTGAGAGAGGAGCGCGTTC 4578
1162 IleLysGlnLeuTyrPdnGlnLysThrArgGlyGlnAspIleAspTyrI 1178
4579 CAGGAGAGAGCGAGAACTCACTGCCAGGTGGAGAACTGAGTAGAA 4628
1178 eSerIleAsnSerAspSerGlnGlyAlaGlyThrArgSerTyrSerTyrA 1195
4629 ACTGCTGACTGTGACCAAGCAGAGAGGTGCAGAGCAAGCTGAAG 4678

1195 rGValValMetGlnThrGlyAspAlaGlnLeuGlnMetArgGlyArgCys 1211
4679 CTGTCCAGGCTCAGGAGGAGGAGCGAGCGGCGGCTTCGAG 4728
1212 SerAlaGlnLysValLeuAlaSerLeuIleLeuArgLeuAlaLeuAl 1228
4729 GCCAGCTGAATGAACTGCAGACCCAGTTGAGCCAGAGAGAGCGAGC 4778
1228 agLThrPheCysLeuAsnCysGlyIleLeuAlaLeuAspGluProThrT 1245
4779 TGAGCACATTAAGCTGCG.....ATGAGAAACCCAAA 4813
1245 hrAsnLeuAspGlyProAsnAlaGlnSer.....LeuAlaAlaLeu 1259
4814 CACATTATGATGCCAAGAGCAGAGCAAGCAAGAGCTGAGAGAGCTG 4863
1260 LeuArgIleMetGlnAlaArgLysGlnGlnLysPheGlnLeuIleVa 1276
4864 CGAGCCTGAGAGCTGCGAGAG.....GAAACCAAGAGCTG.... 4902
1276 IleThrHisAspGlnArgPheAlaHisLeuIleGlyGlnArgGlnLeuA 1293
4903 CGAGCTGAAGCTGAACGCTGGCCATGAGCTACAGAGGCTGGGCTGA 4951
1293 LsGlnLys 1295
4952 AGACCAAG 4959

seq_name: /cgn2_6/ptodata/2/lna/5A_COMB.seq:US-08-467-781-3

seq_documentation_block:
Sequence 3, Application US/08467781
Patent No. 5780596

GENERAL INFORMATION:
APPLICANT: TOBRATILY, GARY
APPLICANT: LINDARD, GRAHAM P
TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA, HURWITZ & THIBEAULT
STREET: 125 HIGH STREET
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02110

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/467,781
FILING DATE: 06-JUN-1995
CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:
NAME: PITCHER ESQ. EDMUND R
REGISTRATION NUMBER: 27,829
REFERENCE/DOCKET NUMBER: MTP-013
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7100
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 6306 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA

FEATURE:
NAME/KEY: CDS
LOCATION: 1..6306
PUBLICATION INFORMATION:

AUTHORS: COMPTON, DUANE A
 AUTHORS: SZILAK, ILVA
 AUTHORS: CLEVELAND, DON W
 TITLE: PRIMARY STRUCTURE OF NDMA, AN INTRANSCULAR
 TITLE: PROTEIN THAT DEFINES A NOVEL PATHWAY FOR
 TITLE: SEGREGATION OF PROTEINS AT MITOSIS
 JOURNAL: J Cell Biol.
 VOLUME: 116
 PAGES: 1395-1408
 DATE: 1992
 US-08-467-781-3

alignment scores:
 Quality: 376.50 Length: 1486
 Ratio: 0.510 Gaps: 62
 Percent Similarity: 49.664 Percent Identity: 19.314

alignment block:
 US-09-538-396-2 x US-08-467-781-3 ..

Align seg 1/1 to: US-08-467-781-3 from: 1 to: 6306

103 GlnlyAlaSerlyMetGluPhelValAlaIleGluSerValLeuGln.. 118
 802 CAGGGCCGACCCCTGAGAGCCAGAGCTTGAGAGCTGCTGCTGACA 851
 119ThrlleAsnProHlsThrGlyGluLysValCys... 129
 852 GAATGAGAGCTTACATGCGGCTGCATGAACCTGAGACAGTGCAGG 901
 130 ..LeuSerTYrArgCysAlaAspMetAspArgGluIleProAlaLeuMet 145
 902 ACCTGAGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 939
 146 GlyAlaSerlyAlaValLeuGluAsnValIlePheValHlsGlnAspG 162
 940AACCACTTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 965
 162 UserAsnTYrProLeuGlnAspProSerThrLeuLysLysPheAsp 179
 966 CCTTCCCTTAAAGTGGGGAGTGGCAGATCATCTGCAGAGCTACAGG 1015
 179 spllePheSer.....AlaThrArgTYrThrLysAla..... 189
 1016 ATGCCCTCAATGAGCTGACGAGAGAGAGAGAGAGAGAGAGAGAG 1065
 190 LeuGluValIleLysLysLeuHlsLysAsp.....GlnMetGlnG 203
 1066 CTAG 1115
 203 uileLysThrPheArgLeuLysLeuGluAsnLeuGlnThrValLysAsp 220
 1116 CAAGAAATGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1156
 220 InAlaHlsLysLeuArgGluAsnIleAlaGlnAspGlnGluLysSer 236
 1157 AACTTTCACAGCTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1206
 237 AlaSerlySer.....GlnMetGluGlnLeuLys 246
 1207 CAG 1256
 246 sGluLysIleCysGlyThrGluArgGluIleLeuGlnMetGluThrSer 263
 1257 GCAG 1306
 263 euAspGluLeuArgArgLeuGlnGlnIleAspIleLys..AlaThr 278
 1307 TAGAGATGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1356
 279 GluArg.....SerThrLeuLeuTh 285

1357 GAGCGGGGCCACTTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1406
 285 rGlnGlnHlsGluLysLeuAlaIleUserGluGluAsnGluAspThra 302
 1407 TGACCTCCAGAGCTCATCTCCACCTCAGCCAGGAGAGAGAGAGAG 1456
 302 sPlGluGluMetGluTYrGlnThrLysPheGluGluArgIleAlaLeu 318
 1457 AGCAGGCTTCCAGGCTCATGGGGCCGGTGGTACCTGCCAGGTGGCT 1506
 319 LeuGluThrLysIleSerLysLeuValArgAspMetAspAspGluAla 335
 1507 CTGACCTCTGAGCTCAACATCAATGCCACCATCAGAGAGAGAGATCA 1556
 335 rTYrSerValLeuSerLysGlnAsnSerGluLeuThrHlsGluIleG 352
 1557 AGAAGCTGCTGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1606
 352 LysLys..LeuGlnAlaGluAlaAspAlaHlsLeuThrMetLysHls 366
 1607 CACAGAGCTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1656
 367GluArgAspSerAspIleLysAs 374
 1657 GTGAGAGAGCTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1706
 374 nileCysTYrLysHlsAsnLeuGluProValProGluHlsProPheThra 391
 1707 GGTAGCGGAGAGAG.....CAGAGAGCAACTAGGC 1735
 391 snAspValAlaMetAsnLeuThrAsnArgIleLysAlaGluSerSer 407
 1736 AG 1785
 408 LeuGluAsn.....AspLeuLeuAspLysLys 416
 1786 TTAAG 1835
 416 sLysSerAsnGluAspGlnLeuAspValLeuTYrPylHlsTYrLeuLys 433
 1836 GAAGGCT.....GCCAAGCTGAGATTCGACAGAGAGAGAGAGAG 1879
 433 LeuAsn.....AlaArgTYrSerGluValAspGluLys 443
 1880 CTATAG 1929
 444IleGlnSerLysIleGluSerMetSerGlyIleLeuArg 456
 1930 GAG 1979
 456 g.....ArgArgLysAspLysGluArgAspAlaGluValG 471
 1980 GACAG 2029
 471 LuLeuSer..LysPheAsnLeuSerArgIleAspGluArgGluArgHls 486
 2030 AGTTGACAGCTGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2079
 487 MetGlnIleGluValGluArgLysThrLeuAlaLeuGluLysArgAsp 503
 2080 GCCCAG.....GAGAAGAGC.. 2094
 503 rAspSerIleLeuSerGlnLysArgThrGluValTYrSerLeuGluLys 520
 2095CAGCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2128
 520 ysIleLysValLeuLeuArgGluLysAspIleIleAsnArgAspAlaAsp 536
 2129 CTTGAAGGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2175
 537 GluArgValLysLeuGluLysLysAspAlaLeuGluSerLysAs 553
 2176GCAGATGCTCCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 2201

553 pLYLeuasnGLuIleValaEnGLuHISLysAspLysIleLysValL 570
2202 TTGATCTCTAGCTGAGAGCAGAGACCCGAAAGCTGTGGAGAGCAT 2251
570 euArgGLyArgAsnProPheGLuLysAsp 579
2252 AGCGGGAACGAAAGAGCTGAGAGAGAGGCGTGGCGCAAGGGCTG 2301
579 579
2302 GAGCTCGATTACTGACAGCTTGGGAGGCCCATCAGCTGAGACTGAAGT 2351
580 .MetLysLysGLuIleasnGLuAlaPhe.....TrpProValAspL 593
2352 CCTGGCGGGAGGCTGGCAGAGGCCATGCTGCCAGCACAGCTGAGA 2401
593 YSGlUTyAsnGLuLeu..... 598
2402 GTGAGCTGAGACAGCTCTCAAGAAGTAGCTGGCGCTGACGGGTAT 2451
599 ..ArgSerLysSerGLuIleaglInGLuIleuLysPheThrGLnSe 614
2452 GAGATAGCCAGCAGAGAGAGGACAGATAGCGCCATGTTTC...CAGA 2498
614 rLysValThrAspAlaArgGLuInLeuThrLysLeuArgArgAspMet. 630
2499 ACAGCTGATGACTTGAAGAGAGAAATGTAGAGAGGCCGCCAGAGACTG 2548
631 ..AspAlaLysArgArgPheLeuAspSerLysLeuInSerIleLeuGLn 646
2549 AGGAGCGAAAGAGAGAAGTGGCAGGCAATGCAACAGCGAGCTCCAG 2598
647 IleSer.....AlaAsnValAspMet 653
2599 ATAAAGCCGCGACAGAAACAATACTAGCTGATCCATGCCCAACTG.... 2643
653 tHeProLysValLeuGLnAspAlaMetAsnLysArgAspGLuInLysA 670
2644 ...GCCAAGACACTCCAGCAGAGTCCAGAGAGAAAGTCAAGGCCAG 2689
670 rGLnLeuAsnPheAlaAsnGLyMetArgLumetLeuAlaPro..... 684
2690 AGCTTGACATGACCTCTCCACTCTCAGAGAAAGATGGTGCACACAGC 2739
685PheGLuHISLeuAlaArgLys..... 691
2740 AAAGAGTGGCCCGCTTGGAGACCTTGTCCGCAAGCAGAGTGAGAGCA 2789
692AsnHISValCysProCysGLuArgAlaPheThr 704
2790 GGAAACAGCTCCCGGAGTACTCAAG.....GAGCTGCGAGGGCAG 2833
704 roAspGLuInAspGLuPheValLysGLnArg...MetGLnAsnSer 719
2834 GAGACAGACAGCCCGAGTGGCTGGAAGAGCAGAGGAGCCAGTTCTGC 2883
720 SerThrAlaGLuArgSerLysAlaLeuAlaMetLysSerSerAsnAlaGL 736
2884 AGCAGACAGCAGCGCTGACGCTATGAGAGCGGAG...GCAGA 2924
736 uAlaLeuPheGLnGLuLeuAspLysLeuArgThrIleTyAspAlaArgY 753
2925 GCAGATGGCGCAATGAGCTGGAACGGCTGCG...G 2956
753 alLysLeuValGLuInLThrIleProLeuAlaGLuLysAsnLeuAsnGLn 769
2957 CCGCGCTGATGAGAGCCAGGCGCAGCAGCAGAGAGAGCGCTGGCAGCAG 3006
770HISLeuAlaAspGLuSerGLnLysAlaGLnAla.. 780
3007 GAAAGGAGAGTGGCGGCTGACCCAGAGCGGCGCTGCCAGAGCTGA 3056
780 780
3057 CTTGCCCTGGAGAAAGCGCCAGAGCAGAGCTTGATGCCGCTGCAGA 3106
781PheAspAspLeuLeuGLyValLeu 788
3107 AGCCCTCAGAGAGAGCGGTGGAGTTCGCACCTCTGCAAGAGGCACTG 3156
789 AlaHISValGLnMetAspArgAspAlaValGLuAlaLeuLeuGLnProth 805
3157 GCTCATGGCCCTGAGGAAAGAGAGCAGAGCAGAGTTGGCCCAACT 3206
805 rAspThrIleAsp.....ArgHISValHISGLuIleGLnGLnL 818
3207 TCGTGTCTGGAGCGAGCCAGATAAAGAGCTGAGAGAACTCGGCAAA 3256
818 euValLysGLuVal.....GLuAspLeuGLuTyAla 828
3257 CCGTAGAGCACTGAAAGAGACAGCTGGCTGAAGAAAGAAAGAGAGCAGCA 3306
829 LeuAspSerSerGLyArgGLyValLysSer..... 838
3307TrpGCGTCAAGAGCCCAATCTGAGCTGTGCGAGAGACA 3347
839LeuGLuGLuIleGLnLeuGLuLeuAsnPheLeu 850
3348 GCCAAGAGCCCGCAAGCTGGAAGCACTGGCGCAGAGGTGAGCAAGCTGG 3397
850 LArg..... 851
3398 AACAGCAATCCAGAAAGCAGAGAGCAGAGCTGACAGCTGGAAACCCAGC 3447
852ThrArgAspThrLeuIleValGL 859
3448 CTGAGAGCTGAGCGGCGCTCCGCGGTAGCGGAGACAGTCTGTGAGAC 3497
859 uValAsp...AspLeuArgAspGLnHISArgMetLeuAsnGLuAspMetS 875
3498 TCTGACAGGCGCAAGTTAGAGAGAAAGCCAGAGAGTGAAGCAACAGTCA 3547
875 er.....SerAlaGLn.....Val 879
3548 GTCCCTTAGCCCTGCGCCCAAGGAGTGGCTCTCCGACCAAGATGA 3597
880 ArgThrHISAsnAlaArgGLuInLysValLysAlaSerSerIleLeuGL 896
3598 CAAGACCAAGCAAGAGCTGAAGATGAGTGAAGCCAGAGTGGCCGCGG 3647
896 uArgPheGLnLysSerGLuInLysLeuValLeuValLeuAlaGLuInLys 913
3648 CCGCAAGAGCTGAGAGGAAATAAGCCATCAGCAGAGCTTGAGAGAGG 3697
913 LuInLeuIleValGLuLysLysLeuLeuGLuInLysLeuAspProLeu 929
3698 AGGTGTCCATCTGAAATCCGACAGTCTGGAAGAGGGGGGAG..... 3741
930 SerLysGLuLysGLuSerLeuLeuGLnGLuTyAsnAlaLeuLysGLnL 946
3742 AGCAAGAGTGAAGCGGCTGTGATGGCGAGTCAAGAAAGAGCCAGAA 3791
946 sLeuAspGLu..... 949
3792 GCTGAGAGAGAGTGCCTCTGTCAGAGCAGAGCAGCAGCAACAGTGC 3841
949 949
3842 CAGAGCTGCAAGAACAGAGCTCTGCTGCGGAGAGAGTGCAGAGCTCC 3891
950GLuTyHISGLnLeuAlaGLuArg.....LysArgGL 960
3892 GGGAGGAGAGCTGGAAGAACAGCGGCTGCTCAAGAAACCTGGCGCAGGA 3941
960 uPheGLnGLuLeuAspAlaLeuGLyArgLeuAsnMetLysIleLysG 977

```

3942 GCTGACCTACAGAGCTGAGCTGCGGAGAGAGCTGCGCAAGATTTGAAG 3991
977 LTYTLEuAspSer....LysLysasnLysLysLysLysLysLys 991
3992 CGTGCGAGAGAAAGTTCTTCAGAAAGAGAGCGCTCTCCACCTCGAG 4041
992 GLYArgHisVal..... 995
4042 CTCGAGCAGACAGCAGAGCCCTGCTGAGTGAAGCTGCTCCAGCTAA 4091
996 ....LeuCyHisSerLysLysLysLysLysLysLysLysLysLys 1011
4092 GCACCTCTGCGAGAGCTGAGCGGAGAGAGCGCTGCGCAAGAGCC 4141
1011 LserAlaLysLysLysLysLysLysLysLysLysLysLysLys 1019
4142 ACCGTGAGAGAGCTGAGAGAGAGAGAGAGCGCTGCGGAGCTGCGGCA 4191
1020 GLULeuLysLysLysLysLysLysLysLysLysLysLysLys 1036
4192 GAGCTGCTGCGGCGCGGAGAGCTGAGAGAGCTGCTCTGCGGCA 4241
1036 ULysLysLysLysLysLysLysLysLysLysLysLysLysLys 1053
4242 GAGGTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4291
1053 LysLysLysLysLysLysLysLysLysLysLysLysLysLys 1069
4292 CCAGCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4330
1070 AlaAspLysArgHisSerLysLysLysLysLysLysLysLysLys 1086
4331 .....AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4349
1086 AAsnArgTrpLysLysLysLysLysLysLysLysLysLysLys 1103
4350 GAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4387
1103 LysLysLysLysLysLysLysLysLysLysLysLysLysLys 1119
4388 AGTTCTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4428
1120 ThrAsnLysLysLysLysLysLysLysLysLysLysLysLys 1133
4429 GTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4478
1133 LysLysLysLysLysLysLysLysLysLysLysLysLysLys 1148
4479 GGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4528
1149 ..ArgPheHisSerMetLys.....MetGluLysLysLysLys 1161
4529 CCAAGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4578
1162 LLeuLysLysLysLysLysLysLysLysLysLysLysLys 1178
4579 CAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4628
1178 AserLysLysLysLysLysLysLysLysLysLysLysLysLys 1195
4629 ACTGCTGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4678
1195 rGValValMetGlnThrGlyAspAlaGluLysLysLysLysLys 1211
4679 CTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4728
1212 SerAlaGlyLysLysLysLysLysLysLysLysLysLysLysLys 1228
4729 GCCAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4778
1228 aGluThrPheCyLeuAsnCyGlyLysLysLysLysLysLysLys 1245

```

```

4779 TGAGCACTATAAGCTCGAG.....ATGAGAGAGAGAGAGAGAG 4813
1245 hAsnLysLysLysLysLysLysLysLysLysLysLysLysLys 1259
4814 CACATTTATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4863
1260 LeuArgLysLysLysLysLysLysLysLysLysLysLysLys 1276
4864 CGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4902
1276 LLeuThrHisLysLysLysLysLysLysLysLysLysLysLys 1293
4903 CGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4951
1293 LysLysLysLysLysLysLysLysLysLysLysLysLysLys 1295
4952 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4959

seq_name: /c9n2_6/prodata/2/lna/5B.COMB.seq:US-08-483-924-3
seq_documentation block:
; Sequence 3, Application US/08483924
; Patent No. 5882876
; GENERAL INFORMATION:
; APPLICANT: TOOKATLY, GARY
; APPLICANT: LIDGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: 125 HIGH STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,924
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESQ, EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6306 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..6306
; PUBLICATION INFORMATION:
; AUTHORS: COMPTON, DUANE A
; AUTHORS: SZILAK, ILYA
; TITLE: PRIMARY STRUCTURE OF NUMA, AN INTRANUCLEAR
; TITLE: PROTEIN THAT DEFINES A NOVEL PATHWAY FOR
; TITLE: SEGREGATION OF PROTEINS AT MITOSIS
; JOURNAL: J. Cell Biol.
; VOLUME: 116
; PAGES: 1395-1408
; DATE: 1992

```

US-08-483-924-3

alignment_scores:	
Quality:	376.50
Ratio:	0.510
Percent Similarity:	49.664
	Length: 1486
	Gaps: 62
Percent Identity:	19.314

```
alignment_block:
  uc-00-538-306-3  uc-08-483-924-3
```

Align seg 1/1 to: US-08-483-924-3 from: 1 to: 6306

103 GlnLysAlaSerLysMetGluPheLysAlaIleGluSerValLeuGln... 118
||| ||||| :|||||
119 ThrIleAsnProHIsThrGlyLysValCys.... 129
||| ||||| :|||
852 GAATAGAGCGCTTACCATGGCGCGCTGCATGAACCCGTAAACCAGTCGCCAGG 901
130 ..LeuSerTyArgCysAlaSpMetAspArgGluIleProAlaLeuMet 145
|||:: :: :|||:
902 ACCGTAAAGACAGAGAAGAGCCAGATGATGCATAAATC..... 939
146 GlyValSerLysAlaValLeuGluAsnValIlePheValHisGlnAspGl 162
940 AACCCAGCTTTCCGAGGAGGAATTGGACA 965
162 UserAsnTrpProLeuGlnAspProSerThrLeuLysLysPheAspa 179
:||||| :|||:
966 CTTTCCTCTTAAGCTGCGGGAGTTTGCCAGTCATCTGCACAGCTTAAGG 1015
179 sPIIepheser.... AlaThrArgTryThrLysAla..... 189
|| :|||:
1016 ATGCCCTCAATGAGCTGACGGAGGAGCACACAGAGGCCACTCAGAGATGG 1065
190 LeuGluValIleLysLysLeuHisLysAsp..... GlnMetGlnGl 203
||||| :|||:
1066 CTAGAGAACAGCAGGCCACCTGAGAACAGAGCTCAGCCAGCCCTCGAGGA 1115
203 uLelysthrPheAdgLeuLysLeuGlnAsnLeuGlnThrValLysasp 220
: ||| :|||:
1116 CAAGAAATGCTCTTAGAGAAAGAACGAAATCTTCAG..... GGA 1156
220 InAlaHisLysLeuArgLysnIleAlaGlnAspGlnGluLysSerAsp 236
:||||| :|||:
1157 AACTTTCACAGCTGAGAGAACACTGTCTCCAGCTGCAGATAAACCCACC 1206
237 AlaSerLysSer..... GlnMetGluGlnLeuLy 246
:||||| :|||:
1207 CAGAGAAAGCGGAGAGTGTGCTGGATGATCTTCCAGCTGGAACCTTGA 1256
246 sGluLysIleCysGlyThrGluArgGluIleLeuGlnImetGluInhrserL 263
:||||| :|||:
1257 GCAGAGGCGAGCCACTTTCGTCAAAACACACAGAGCTCCAAAGCCAGG 1306
263 euAspLysLeuArgTrgLeuGlnIlyGlnIleAspIleLys... Alathr 278
:||||| :|||:
1307 TGAAGTGTCTGGAGACTGAGCGAGGCCACAGAAACCAAGCTCTTGCT 1356
279 GluArg..... SerThrLeuLeuTh 285
||||| :|||:
1357 GACGGGGGCCACTTCGAAAGAAAAGCAGCAGCTGTCTAGACGTGATCAC 1406
285 rGlnGlnHisGlyLysLeuAlaAlaLeuSerGluLysnGlnLyspTrpA 302
:||||| :|||:
1407 TGACCTGCAGAGAGCTCCATCTCCAACTCTCAGCCAGGACCAAGAAAGCTGG 1456
302 spLIuLysLeuMetGluTrpGlnIlnThrLysPheGluGluArgIleLeu 318
:||||| :|||:
1457 AGCAGGCGCTCCACAGGCTCATGTGGGGCCGGTTTACTGCTCCAGAGTGGCTCT 1506

[illegible]

[illegible]

```
seq_documentation_block:
```

; Patent No. 5783403

APPLICANT: YOU

TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	1																																																																																								

ADDRESSEE: TESTA HURWITZ & THIBEAULT

CITY: BOSTON

STATE: MA
COUNTRY: USA

21F. 02103
COMPUTER READABLE FORM:
;

COMPUTER: IBM PC compatible

```

; OPERATING SYSTEM: PC DOS/MS DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

```

;;
; COMMENT AFFILIATION DATA:
; APPLICATION NUMBER: US/08/195,487

FILED DATE: 435
CLASSIFICATION: 435

```

; PRIOR APPLICATION DATA.
; APPLICATION NUMBER: US/07/901,701
;

```

ATTORNEY/AGENT INFORMATION:

NAME: FICHER ESQ, EDMOND A
REGISTRATION NUMBER: 27,829

REFERENCE/DOCUMENT NUMBER: [REDACTED]
TELECOMMUNICATION INFORMATION:

TELEPHONE: 017/248 7100
TELEFAX: 617/248-7100

```

; INFORMATION FOR SEQ ID NO: 1
;
; SEQUENCE CHARACTERISTICS:

```

```

; LENGTH: 800 base pairs
; TYPE: nucleic acid

```

```

; SIGNEDNESS: single
;
; TOPOLOGY: linear
;

```

```

; MOLECULE 1FE: CDNA
; FEATURE:

```

NAME/REF: CDS
LOCATION: 11.6306

;; PUBLICATION INFORMATION.
;;
;; AUTHORS: COMPTON, DUANE A

AUTHORS: CLEVELAND, DON W

TITLE: PROTEIN THAT DEFINES A NOVEL

JOURNAL: JOURNAL OF CELL BIOLOGY

VOLUME: 118
PAGES: 1395-1408

DATE: MAR 1972
US-08-195-487-3

alignment_scores: 374.50
quality: 8.70

Percent Similarity: 49.596 Percent Identity: 49.596

Alignment block:

US-09-538-396-2 x US-08-195-487-3

Align seq 1/1 to: US-08-195-487-3 from: 1 to: 6306

```

103 GluLysAlaSerLysMetGluPheLysAlaIleGluSerValLeuGln.. 118
104 ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
802 CAGCGCGCCAGCCACCTGAGAGCCCAAGAGAGCTTGAAGAGCTGGTGACAA 851
119 .....ThrlLeaProHisThrGlyLysValCys... 129
120 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
852 GAATGAGAGCTTACCATCGCGGCTGATGAACCCCTGAAGCAGTGCAGG 901
130 ..LeuSerTyrArgCysAlaAspMetSparGluIleProAlaLeuMet 145
131 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
902 ACCGTGAAGACAGAGAGAGCCAGATGGATGCAAAATC..... 939
146 GlyValSerLysAlaValLeuGluAsnValIlePheValHisGlnAspG1 162
147 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
940 .....AACCACTTTCGAGAGAGAGATGAGCA 965
162 UserAsnTyrProLeuGlnAspProSerThrLeuLysLysLysPheAspA 179
163 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
966 CCTTCTTAACTGCGGAGTTGGCACTCATCTGCAGCAGCTACAG 1015
179 sPrlPheSer....AlaThrArgTyrThrLysAla..... 189
180 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1016 ATGCCCTCAATGAGCTGAGAGAGAGACAGCAAGGCCACTCAGAGAGTGG 1065
190 LeuGluValIleLysLysLeuHisLysAsp.....GlnMetGlnG1 203
191 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1066 CTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1115
203 uileLysThrPheArgLeuLysLeuGlnAsnLeuGlnThrValLysAspG 220
204 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1116 CAAGAAATGCTTGAAGAGAAAGAAATCCTTCA.....GAA 1156
220 LnaIaHisLysLeuArgGluAsnIleAlaGlnAspGlnGluLysSerAsp 236
221 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1157 AACCTTCACAGCTGGAAGAACACTTGTCCAGCTGCAGAGATAACCCACC 1206
237 AlaserLysSer.....GlnMetGlnGlnLeuLys 246
238 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1207 CAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1256
246 sGluLysIleCysGlyThrGluArgGluIleLeuGlnMetGluThrSerL 263
247 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1257 GCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1306
263 euAspGluLeuArgArgLeuGlnGlnIleAspIleLys...AlaThr 278
264 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1307 TAGAGATGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1356
279 GluArg.....SerThrLeuLeuThr 285
280 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1357 GAGCGGGGCCACTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1406
285 rGlnGlnHisGluLysLeuAlaLeuSerGluLysGlnLysAspThrA 302
286 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1407 TGACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1456
302 sPrlGluLysLeuMetGluTyrPheGlnThrLysPheGluGluArgIleAlaLeu 318
303 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1457 AGCGAGGCTCCAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1506
319 LeuGluThrLysIleSerLysLeuValArgAspMetAspArgLysAla 335
320 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1507 CTGACCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1556
335 rTyrSerSerValLeuSerLysGlnAsnSerGluLeuThrHisGluIleG 352
336 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1557 AGAAGCTGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1606

```

```

352 LysLys...LeuGlnAlaGluAlaAspAlaHisLeuThrMetLysHis... 366
353 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1607 CACAGACCTTCACACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1656
367 .....GluArgAspSerAspIleLysAs 374
368 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1657 GTGAGCAGAGTAAGCACTAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 1706
374 nileCysThrLysHisAsnLeuGluProValProGlnHisProPheThrA 391
375 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1707 GTAGCGGAGAGAG.....CAGAGAGCAATGAGC 1735
391 snAspValAlaMetAsnLeuThrAsnArgIleLysAlaArgLeuSerSer 407
392 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1736 AGAGACATGCCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1785
408 LeuGluAsn.....AspLeuLeuAspLysLys 416
409 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1786 TTAAGGAGAGCGGAGTGGCTTCAGACAGCTGAGAGAGAGAGAGAGAG 1835
416 slySerAsnGlnAspGlnLeuAspValLeuTyrPheHisTyrLeuLysI 433
417 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1836 GAAGGCT.....GCCAAGCTGGAGATTCTGCAGCAGACCACTTCAGGTGG 1879
433 LeuAsn.....AlaArgTyrSerGluValAspGlyGln... 443
434 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1880 CTATAGTAAGCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1929
444 .....IleGlnSerLysIleGluSerMetSerGlyLeuLys 456
445 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1930 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1979
456 g.....ArgArgLysAspLysGluLysGluArgAspAlaIleGluValG 471
457 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1980 GACAGCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2029
471 LuLeuSer...LysPheAsnLeuSerArgIleAspArgGluArgHis 486
472 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2030 AGTTGACAGCTGGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2079
487 MetGlnIleGluValGluArgLysThrLeuAlaLeuGluLysArgAspTy 503
488 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2080 GCCCAG.....GAGAGAGAG... 2094
503 rAspSerIleLeuSerGlnLysArgThrGluValTyrSerLeuGlnL 520
504 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2095 .....CAGCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2128
520 ysiLeuValLeuLeuArgGluLysAspIleLeuAsnArgAsnAlaAsp 536
521 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2129 CTTGAAGAGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2175
537 GluArgValLysLeuGlyLeuLysLysAspAlaLeuGluSerLysAs 553
538 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2176 .....GCAGATGCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2201
553 PLYLeuAsnGluIleValAsnGlnHisLysAspLysIleLysIleVal 570
554 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2202 TTGTATCTGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2251
570 euArgGlyArgAsnProPheGluLysAsp..... 579
571 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2252 AGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2301
579 ..... 579
2302 GAGGCTGATTAATCAGAGCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAG 2351
580 MetLysLysGluIleAsnGlnAlaPhe.....TyrProValAspL 593
581 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2352 CTGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2401

```

```

593 ysgIuTYrAsnGluLeu..... 598
594 .....
2402 GTGAGTGTGACACAGCTCTCAAGAAAGAGTGCCTGGCGTGAAGGATAT 2451
599 ..ArgSerIysSerGlnGluAlaGluGlnLeuIleuLysPheThrGlnSe 614
2452 GAGATACACCAAGAGAGAGGACACAGTATGGCCGCAATGTC...CAGGA 2498
614 rIyVaIThrAspAlaArgGluGlnLeuThrIleuLysArgArgAspMet. 630
2499 ACAGTGTGACTTGTGAAGGAGGAATGTGAGAAAGGCCCGCAGAGGAGCTGC 2548
631 ..AspAlaLysArgArgPheLeuAspSerIysLeuGlnSerIleLeuGln 646
2549 AGGAGGCAAAAGAGAGAGTGGCAGGATAGAAATCCACAGCAGGAGCTCCAG 2598
647 IleSer.....AlaAsnValAspMet 653
2599 ATAAAGCCGCGCAGCAGAACAACTAGCTGAGCTCCATGCCAACCTG.... 2643
653 rPheProLysValLeuGlnAspAlaMetAsnLysArgAspGluGlnLysA 670
2644 ....GCCAGACCACTCCAGAGCTCCAAAGAGAGAGTCAAGGCCACAGA 2689
670 rIleuGlnAsnPheAlaAsnGlnMetArgGluMetLeuAlaPro..... 684
2690 AGCTTGACAGATGACTCTCCACTGTGCAGAAAGATGGCTGCCACACAGC 2739
685 .....PheGlnIleuAlaArgLys..... 691
2740 AAAGAGTGGCCGCTGTGAGACCTTGGTCCGCAAGGCGAGTGAAGCAGCA 2789
692 .....AsnHisValLysProCysGluArgAlaPheThrP 704
2790 GGAACAGCCTCCCGGAGATTAGTCAAG.....GAGCCTGGAGGCGAG 2833
704 rAspGluGlnAspGluPheValLysGlnArg...MetGlnAsnSer 719
2834 GAGACAGACGCCCGAGTGGCTGGAGAGCAACAGGAGGCCAGTCTGC 2883
720 SerThrAlaGluArgSerLysAlaLeuAlaMetGluSerSerAsnAlaG 736
2884 ACCACACAGCAGCGCTGCAGCTATGAGCGGGAG.....GCAGA 2924
736 uAlaLeuPheGlnGlnLeuAspLysLeuArgThrIleTYrAspAlaTYrV 753
2925 GCAGATGGCAATGAGCTGGAACGCGCTGCGG.....G 2956
753 aLlysLeuValGluGluThrIleProLeuAlaGluLysAsnLeuAsnGln 769
2957 CGCGCTGATGAGAGCGGAGGAGCAGCAGCAGAGGAGCGTGGCGCAGAG 3006
770 .....HisLeuAlaAspGluSerGlnLysAlaGlnAla... 780
3007 GAAAGGAGAGTGGCGCGCTGACCCAGAGCGGCGCTGCCAGGCTGA 3056
780 ..... 780
3057 CCTTGCCTGAGAGAGCGGCGCAGACAGACTTGAGATGCGGCTGCAGAG 3106
781 .....PheAspAspLeuLeuGlyValLeu 788
3107 ACGCCCTCAAGCAGCAGCGTGTGAGTTCCTACCTGCAGAGGCGACTG 3156
789 AlaHisValGlnMetAspArgAspAlaValAlaLeuLeuGlnProTh 805
3157 GCTCATGCCCTGACGAAAGAGAGCAGACAGAGATGGCCAAAGCT 3206
805 rAspThrIleAsp.....ArgHisValHisGluIleGlnGlnL 818
3207 TCGTGTCTGTGAGCAGACCCAGATAAAGAGCTGAGAGAACTTCGGGCAA 3256
818 euValLysGluVal.....GluAspLeuGluTYrAla 828

```

```

3257 CCGTGAACCACTGAAGAGAACAGCTGCTTAAGAAAGAAAGAGACAGCA 3306
829 LeuAspSerSerGlyArgGlyValLysSer..... 838
3307 .....TCTGGCTCAGAGCCCAACTGAGAGCTGTGCGCAGAGCAGA 3347
839 .....LeuGluGluIleGlnLeuGlnLeuAsnPheLeuG 850
3348 GCCAACAGAGCCCAAGCTGGAAGCACTGCGGGCAGAGGTGAGCAAGCTGG 3397
850 LnaArg..... 851
3398 AACAGCAATGCCAGAGCAGACAGCAGAGCTGACAGCTCGGAACGACGC 3447
852 .....ThrArgAspThrIleuIleValG 859
3448 CTGAGAGCTGACGGGCTCCCGGCTGAGCGGAGCACTGCTCTGGAGAC 3497
859 uValAsp...AspLeuArgAspGlnHisArgMetLeuAsnGluAspMetS 875
3499 TGTGAGGGCCAGTTAAGAGAGAGCCCGAGAGCTTAGGGCAGACTCAGA 3547
875 er.....SerAlaGln.....Val 879
3548 GTGCCTTACCTCGGCCCAAGCGAGATTGGCTGCTCCGACCAAGTA 3597
880 ArgThrHisAsnAlaArgGluGluLysValLysAlaSerSerIleLeuG 896
3598 CAAGACCAAGCAGAGCTGAAGATGATGAGAGGCCAGCTGAGGCCGGGG 3647
896 uArgPheGlnLysSerGluGluGluLeuValLeuLeuAlaGluGluLysG 913
3648 CCGGCAAGAGGCTGAGAGAGAAATAGCTCATCAGACACTGGAGAGAG 3697
913 GlnGlnLeuValGluLysLysLeuLeuGlnGluSerLeuAspProLeu 929
3698 AGCTGTCCATCCTGATCCCGAGCTCTGAGAAAGAGGGAG..... 3741
930 SerLysGluLysGluSerLeuLeuGlnGluTYrAsnAlaLeuLysGlnL 946
3742 AGCAAGAGATTGAAGCGCTGTGATGCCGAGTCAAGAAAGCCAGAA 3791
946 sLeuAspGlu..... 949
3792 GCTGGAGAGAGAGTGGCGCTGTCAGCAGAGACAGCCAGCAAGCTGC 3841
949 ..... 949
3842 CAGAGCTGCAGAACGAGCTGTCTGCGGAGAGAGTGCAGAGCCTCC 3891
950 .....GluTYrHisGlnLeuAlaGluArg...LysArgG 960
3892 GGGAGGAGAGCTGAGAAACAGCGGTGGCTTCAAGAACTCTGCGCAGAG 3941
960 uPheGlnGlnIleuAspAlaLeuGlyArgLeuAsnMetLysIleLysG 977
3942 GCTGACCTCAGAGCTGAGCGTGGAGAGAGCTGGGCGCAAGATTGAGG 3991
977 LYTyrLeuAspSer.....LysLysAsnGluLysLeuLysGluLeuGln 991
3992 CGTGGCAGAGAGATTCTTCCAGAAAGAGCGCCTCTCCACCTGCAG 4041
992 GlyArgHisVal..... 995
4042 CTCGAGACACACAGCAGAGCCCTGTGTGATGAGCTGTGCCAGCTAA 4091
996 .....LeuCysHisSerGlnLeuGlnSerCysMetAlaLysGlnArgT 1011
4092 GCACCTTGCCACAGCTGAGGCGGAGCGGCGCGGAGAAAGGCC 4141
1011 leSerAlaGluLeuAsnLysSerLys..... 1019

```


4142 ACCGTGAGAGCTGAGCAGAGCAGAGCGCGCTGGGGAGCTCGGGCA 4191
 1020 GlnLeuLeuGlnGlnGlnGlnLeuLysArgAsnIleAspAspAsnLe 1036
 4192 GAGCTGCTGGGGCCGCGAGCGAGCTGGGGAGCTGATTCCTGCGGCA 4241
 1036 UlystYArgLysThrLysAlaAspValGlnGlnLeuThrArgAspIleG 1053
 4242 GAGGTGGCAGAGCAGAGCAGACGACGCTGAGAGCTGGGGCAGAGAGG 4291
 1053 IuSerLeuGlnGlnArgLeuLeuSerIleGlySerLeuSerAlaIleGlu 1069
 4292 CCAGCTTGGAGAGCAGCAGCTGAGCTG..... 4320
 1070 AlaAspLeuLysArgHisSerGlnGlnLysGlnArgLeuAsnSerGluPh 1086
 4321AAGAAAGCGCATGGC.....CTGCTGGCAGAGGA 4349
 1086 eAsnArgTrpGlnGlnThrLeuSerValTyrGlnSerAsnIleSerLysH 1103
 4350 GAACTGGGGGGCTGGGTAG.....CGGGCAACCTTGGCCGGC 4387
 1103 IsLysGlnGlnLeuLysLeuSerGlnTyrLysAspIleGlnLysArgTyr 1119
 4388 AGTTCTGGAAGTGGAGTTGGACCGCCGGGA.....AAGTAT 4428
 1120 ThrAsnGlnPheLeuGlnLeuLysThr.....ThrgluMetAlaAs 1133
 4429 GTCCAAAGAGTGGCGAGCGCTGATGCTGAGACCGCTGCTGTA 4478
 1133 nLysAspLeuAspArgTyrTyrThrAlaLeuAspLysAlaLeuMet... 1148
 4479 GGTGCGAGAGAGCAGACAGCACTGCCGGGAGCTGGAGTGTAGCTG 4528
 1149 ..ArgPheHisSerMetLys.....MetGlnGlnIleAsnLysIle 1161
 4529 CCAAGTATGAGGCTGCCAAGTCAAGGCTCGAGAGAGAGCAGCGTTC 4578
 1162 IleLysGlnLeuTrpGlnGlnThrTyrArgGlnLysAspIleAspTyrI 1178
 4579 CAGGAAGAGAGAGCAAACTCATCTGCCAGGTGGAGAGAACTGAGTAAGA 4628
 1178 eSerIleAsnSerAspSerGlnGlnLysAlaGlyThrArgSerTyrSerTyr 1195
 4629 ACTGCTGACTCTGACCAAGCAGCAGAGTGCACGACGAGAGCTGAAGG 4678
 1195 rGValIValMetGlnThrGlyAspAlaGlnLeuGlnMetArgGlyArgCys 1211
 4679 CTGTCGAGGCTCAGAGGAGCGAGAGCCGAGAGGCGCCACGCTTCAG 4728
 1212 SerAlaGlnGlnLysValLeuAlaSerLeuIleIleArgLeuAlaLeuAl 1228
 4729 GCCCAGCTGATGAACTGCAAGCCAGCTTGAAGCCAGAGAGGAGCAGC 4778
 1228 aGluThrPheLysLeuAsnGlyIleLeuAlaLeuAspLysProThrT 1245
 4779 TGAGCACTATTAAGCTGCG.....ATGGAGAAACCCAAA 4813
 1245 hAsnLeuAspGlyProAsnAlaGlnSer.....LeuAlaAlaLeu 1259
 4814 CACATTAATGATCCAGAGAGCAGAGCAAGACCTGACAGACAGAGTGTG 4863
 1260 LeuArgIleMetGlnAlaArgLysGlnLysAsnPheGlnLeuIleVal 1276
 4864 CGGAGCTGAGAGCTGACAGAG.....GAAACAAAGAGCTG.... 4902
 1276 IleThrHisAspGluArgPheAlaHisLeuIleGlyLysArgGlnLeu 1293
 4903 ..CGAGCTAACTGAACGCTGGCCATGAGCTACAGCAGGCTGGGCTGA 4951
 1293 IagLysLys 1295
 4952 AGACCAAG 4959

seq_name: /cgn2_6/ptodata/2/lna/PCTUS_COMB.seq:PCT-US93-06160-3
 seq_documentation_block:
 ; Sequence 3, Application PC/TUS9306160
 ; GENERAL INFORMATION:
 ; APPLICANT:
 ; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
 ; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
 ; NUMBER OF SEQUENCES: 6
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: TESTA HURWITZ & THIBEAULT
 ; STREET: 53 STATE STREET
 ; CITY: BOSTON
 ; STATE: MA USA
 ; COUNTRY: USA
 ; ZIP: 02109
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US93/06160
 ; FILING DATE: 19930621
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: PITCHER ESQ, EDMUND R
 ; REGISTRATION NUMBER: 27,829
 ; REFERENCE/DOCKET NUMBER: MTP-013
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 617/248-7000
 ; TELEFAX: 617/248-7100
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 6306 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 1..6306
 ; PUBLICATION INFORMATION:
 ; AUTHORS: COMPTON, DUANE A
 ; AUTHORS: SZILAK, ILIYA
 ; TITLE: PRIMARY STRUCTURE OF OF NUMA, AN INTRANUCLEAR
 ; TITLE: PROTEIN THAT DEFINES A NOVEL PATHWAY FOR
 ; TITLE: SEGREGATION OF PROTEINS AT MITOSIS
 ; JOURNAL: JOURNAL OF CELL BIOLOGY
 ; VOLUME: 116
 ; PAGES: 1395-1408
 ; DATE: MAR-1992
 ; PCT-US93-06160-3
 alignment_scores:
 ; Quality: 374.50 Length: 1486
 ; Ratio: 0.508 Gaps: 62
 ; Percent Similarity: 49.596 Percent Identity: 19.314
 alignment_block:
 US-09-538-396-2 x PCT-US93-06160-3 ..
 Align seg 1/1 to: PCT-US93-06160-3 from: 1 to: 6306
 103 GlnLysAlaSerLysMetGluPheLysAlaIleGlnSerValLeuGln.. 118
 802 CAGCGGCGCAGCCAGCTGAGCTGAGAGAGAGCTGAGAGCTGCGTGA 851
 119ThrIleAsnProHisThrGlyLysValCys.... 129

852 GAATGAGACCTTACCATGGCGCTGCATGAAACCTGAAAGCAGTGCAGG 901
130 ..LeuSerTyrArgCysAlaAspMetAspArgLulIleProAlaLeuMet 145
902 ACCTGAAGACAGAGAAAGACAGATGATCCGAAATC..... 939
146 GlyValSerLysAlaValLeuGluAsnValIlePheValHisGlnAspGlu 162
940AACCAAGCTTTCGAGAGAGAAATGGAGAA 965
152 userAsnTrpProLeuGlnAspProSerThrLeuLysLysLysPheAspA 179
966 CATTTCCTTTAAGCTGCGGAGTTCGCCAGTATCTGCAGACGATACAGG 1015
179 spIlePheSer.....AlaThrArgTyrThrLysAla..... 189
1016 ATGCCCTCAATAGCTGACGGAGAGACAGACAGAGCCACTCAGAGAGTGG 1065
190 LeuGluValIleLysLysLeuHisLysAsp.....GlnMetGlnGlu 203
1066 CTAGAGAACACAGCCCGCTGAGAGAGAGAGCTCAGCGCAGCCCTGCAGGA 1115
203 ulleLysThrPheArgLeuLysLeuGluAsnLeuGlnThrValLysAspG 220
1116 CAGAAATGCTTGAAGAGAGAAAGAAATCCTTCAG.....GGA 1156
220 InAlaHisLysLeuArgGluAsnIleAlaGlnAspGlnLysSerAsp 236
1157 AACTTTCACAGCTGGAGAAACACTTGTCCAGCTGCAGAGATAACCCACCC 1206
237 AlaserLysSer.....GlnMetGlnGluLeu 246
1207 CAGAGAAAGCGGAGGTGCTGGGTATGTCCTGACAGCTGAAACCTTGAA 1256
246 sGluLysLecysGlyThrLugArgLulIleLeuGlnMetGluThrSerL 263
1257 GCAGAGGCGAGCACTCTGCTGCAGAAACACACACAGCTCCAAAGCCAGGG 1306
263 euAspLulLeuArgArgLeuGlnGlyGlnIleAspIleLys.....AlaThr 278
1307 TAGAGATGTGGAGACTGAGCGAGCCAGACAGAAACCAAGCTGCTTGT 1356
279 GluArg.....SerThrLeuLeu 285
1357 GAGCGGCGCACTTGCAGAAAGAAAGACAGACAGCTGTACCTGATCAC 1406
285 rGlnGlnHisGluLysLeuAlaAlaLeuSerGluGluAsnGluAspThra 302
1407 TGACCTGCAGAGCTCCATCTCCAACTCAGCCAGGCCAAGAAAGAGCTGG 1456
302 spGluGluLeuMetGluTrpGlnThrLysPheGluGluArgIleAlaLeu 318
1457 AGCAGGCTCCCGAGCTCATGGGCGCGGTGACTGCCAGGTGGCTCT 1506
319 LeuGluThrLysIleSerLysLeuValArgAspMetAspAspLulase 335
1507 CTGACCTGAGCTCACACACATCAATGCCACCATCCACAGACAGATCA 1556
335 rTyrSerSerValLeuSerLysGlnAsnSerGluLeuThrHisGluLeu 352
1557 AGAACTGGCTGGCTGAAAGCAGAGCCAAAGAAAGAGAGCCCGAGTAG 1606
352 Lys.....LeuGlnAlaGluAlaAspAlaHisLeuThrMetLysHis... 366
1607 CACAGACCCCTCCAAACAGAAACAGAGCCCTCCAGGCTCCGCGACAG 1656
367GluArgAspSerAspIleLysAs 374
1657 GTGAGACAGCTAAGACAGTACCTGAAGACAGAGAGCAGCAGTTGAAGA 1706
374 nIleCysThrLysHisAsnLeuGluProValProGluHisProPheHis 391
1707 GGTAGCGGAGAG.....CAGGAGGCAACTAGGC 1735
391 snAspValAlaMetAsnLeuThrAsnArgIleLysAlaArgLeuSerSer 407
1736 AGACCAAGCCCGACAGCACTGGCCACTGCTGCAGAGAGAGAGGCCCTCC 1785
408 LeuGluAsn.....AspLeuLeuAspLysLys 416
1786 TTAAGGAGAGCGGATGCGGCTCTCAAGCAGAGCTGAGAGCAGTGAAGA 1835
416 sLysSerAsnGluAspGlnLeuAspValLeuTrpLysHisTyrLeuLysI 433
1836 GAAGCT.....GCCAAGCTGAGATTCCTGAGAGAGAACTTCAGGTGG 1879
433 leAsn.....AlaArgTyrSerGluValAspLysGln... 443
1880 CTAAATGAACCCCGGACACTGCCAGACTCTCAGTGCACAGGCCCAAGCGG 1929
444IleGlnSerLysIleGlnSerMetSerLysLysLeu 456
1930 GAGAAAGCAGAGCTGACCGCGAAGGTGAGAGAACTCCAGCGCTGTGTGA 1979
456 g.....ArgArgLysAspLysGluLysGluArgAspAlaAlaGluValG 471
1980 GACAGCCCGCCAGAAACACATGAGAGCCAGGCCCGCAGGTTGCAGAGCTG 2029
471 lueSer...LysPheAsnLeuSerArgIleAspGluArgLulArgHis 486
2030 AGTGTGAGCTGGGTGTGTGACAGCAAAAGGCACTGAGAAAGAAAGGTG 2079
487 MetGlnIleGluValGluArgLysThrLeuAlaLeuGluArgLysArg 503
2080 GCCCAG.....GAGAAAGAC... 2094
503 rasPserIleIleSerGlnLysArgThrLulValLysSerLeuGluGlnL 520
2095CACTCCAGAGAGAGCTCCAGGCCCTCAAGAGCT 2128
520 yalIleLysValLeuLeuArgGluLysAspIleIleAsnArgAsnAlaAsp 536
2129 CTTGAGAGCTGACCAAGGCGAGCTTGAAGAGAGAGAGAGCGAGCT... 2175
537 GluArgValLysLeuLysLysLysAlaLeuGlnSerSerLysAs 553
2176GCAGATCCCTGGAAGAGAGAGAGCGG 2201
553 pLysLeuAsnGluIleValAsnGlnHisLysAspLysIleLysLysValL 570
2202 TTGATCTGTGAGCTGAAGCAGAGACCGAGAGCTGTGGAGACAGATTA 2251
570 euArgGluArgAsnProPheGluLysAsp..... 579
2252 AGCGGAAACGAAAGAGCTGGAAGAAAGAGGCTGGCGCAAGGCGCTG 2301
579 579
2302 GAGCTCGATTACTGCAGCTGGGAGGCCCATCAGGCTGAGCTGAAGT 2351
580 .MetLysLysGluIleAsnGlnAlaPhe.....TrpProValAspL 593
2352 CCGCGCGGAGAGCTGCGAGAGCCATGGCTGCCAGACACAGACAGTAGA 2401
593 ysgLulArgAsnGluLeu..... 598
2402 GTGAGTGTGAGAGCTGTCAAGAAAGTACCTGCGCTGGCGAGCGGTAT 2451
599 ...ArgSerLysSerGlnGluAlaGluGlnGluLeuLysPheThrGlnse 614
2452 GAGATAGCCAGCAAGAGAGAGCAGTAAGGGCGCATGTTC...CAGGA 2498
614 LysValThrAspAlaArgGluGlnLeuThrLysLysLeuArgArgMet. 630
2499 ACAGCTGATGACTTTGAAGAGAGAAATGTGAAGAGGCCCGCCAGAGAGCTGC 2548

```

631 ..AspAlaIySArgArgPheLeuAspSerIySLeuGlnSerIleLeuGln 646
      ::::::::::::::::::::
2549 AGGAGCGAAAGAGAGAGTGGCAGCAGCATAGATCCACAGCAGACCTCCAG 2598
      ::::::::::::::::::::
647 ILIser.....AlaAsnValAspMe 653
      ::::::::::::::::::::
2599 ATTAAGCCGGCAGCAACAACTAGCTGAGCTCCATGCCAACCTG..... 2643
      ::::::::::::::::::::
653 tPheProIySValLeuGlnAspAlaMetAsnIySArgAspIleGlnIySArg 670
      ::::::::::::::::::::
2644 ...GCCAGACACTCCAGCAGCTCCAAAGAGAAGTACAGGGCCAGCA 2689
      ::::::::::::::::::::
670 tYLeuGlnAsnPheAlaAsnIyMetArgIleMetLeuAlaPro..... 684
      ::::::::::::::::::::
2690 AGCTTGACATGACCTCTCCACTCTGCGAGAAAGATGGCTGCCACAGC 2739
      ::::::::::::::::::::
685 .....PheGlnHisLeuAlaArgIyS..... 691
      ::::::::::::::::::::
2740 AAAGAGTGGCCCGCTGGAGACCTGGTCCGCAAGGCAGGTGACAGCA 2789
      ::::::::::::::::::::
692 .....AsnHisValIySProCySArgIySArgAlaPheThrP 704
      ::::::::::::::::::::
2790 GGAACACACCTCCCGGAGGTAGTCAAG.....GAGCCTGCGAGGGCAG 2833
      ::::::::::::::::::::
704 tOAspGlnIySArgIySValIySArgIySArg...MetGlnAsnSer 719
      ::::::::::::::::::::
2834 GAGACAGACAGCCGAGTGGTGAAGAGCAACAGGACGCCAGTTCTGC 2883
      ::::::::::::::::::::
720 SerThrAlaGlnArgSerIyAlaLeuAlaMetGlnSerSerAsnAlaGln 736
      ::::::::::::::::::::
2884 AGCACACAGCGCAGCTGACGCTATGAGCGGGAG.....GCAG 2924
      ::::::::::::::::::::
736 uAlaLeuPheGlnIleuAspIySLeuArgThrIleTyAspAlaTyv 753
      ::::::::::::::::::::
2925 GCAGATGGGCAATGAGCTGGAAAGGCTCCG.....G 2956
      ::::::::::::::::::::
753 aIlyLeuValIleGlnIyThrIleProLeuAlaGlnIySAsnLeuGln 769
      ::::::::::::::::::::
2957 CCGGCTGTGAGAGAGCCAGGGGAGCAGCAGAGAGAGCTGGCAGCAG 3006
      ::::::::::::::::::::
770 .....HisLeuAlaAspIleuSerGlnIySAlaGlnAla... 780
      ::::::::::::::::::::
3007 GAAAGGAGGTGGCGGCTGACCCAGAGCGGGCCGCTGCCAGGCTGA 3056
      ::::::::::::::::::::
780 ..... 780
      ::::::::::::::::::::
3057 CCTTGGCCCTGGAGAAAGGCGCCAGACAGCTGAGATCGGCTGCAGA 3106
      ::::::::::::::::::::
781 .....PheAspAspLeuLeuGlnIyValLeu 788
      ::::::::::::::::::::
3107 AGCCCTCAAGCAGCAGCTGTGAGATTGCTACCTCCAGAGAGCAGCTG 3156
      ::::::::::::::::::::
789 AlaHisValIleMetAspArgAspAlaValIleuAlaLeuGlnProh 805
      ::::::::::::::::::::
3157 GCTCATGCGCTTACGGAAGAAAGAGCAGCAGAGAGTGGCCAGCT 3206
      ::::::::::::::::::::
805 tAspThrIleAsp.....ArgHisValHisGlnIleGlnIle 818
      ::::::::::::::::::::
3207 TGTGGTCTGGAGGAGCCCAATTAAGAGCTGAGAGAACTTGGCAAA 3256
      ::::::::::::::::::::
818 euValIySArgIyVal.....GlnAspLeuGlnIyTyAla 828
      ::::::::::::::::::::
3257 CCGTGAAGCACTGAAGAGAAAGAGCTGTAAAGAAAGAAAGAGCAGCA 3306
      ::::::::::::::::::::
829 LeuAspSerSerGlnArgGlnIyValIySArg..... 838
      ::::::::::::::::::::
3307 .....TCTGGCTCAGAGAGCCCAATCTGAGGCTGCTGCCAGCAGCA 3347
      ::::::::::::::::::::
839 .....LeuGlnIyIleGlnIleuGlnIyLeuAsnPheLeu 850
      ::::::::::::::::::::
3348 GCCAAGAGCCCAAGCTGAGAGCACTCGCGGAGAGGTGAGCAAGCTGG 3397
      ::::::::::::::::::::
850 InArg..... 851
      ::::::::::::::::::::

```

```

3398 AACAGCAATGCCAGAAACAGAGAGCAGGCTGACAGCTGAGACGAC 3447
      ::::::::::::::::::::
852 .....ThrArgAspThrIleuIleVal 859
      ::::::::::::::::::::
3448 CTCGAGGCTGAGCGGCTCCCGGCTGAGCGGAGCAGTGTCTGGAGAC 3497
      ::::::::::::::::::::
859 uValAsp...AspLeuArgAspGlnHisArgMetLeuAsnIySAspMetS 875
      ::::::::::::::::::::
3498 TCTGACAGGCGCATGAGAGAGAAAGCCCAAGACTTAGGCGACAGTCA 3547
      ::::::::::::::::::::
875 er.....SerAlaGln.....Val 879
      ::::::::::::::::::::
3548 GTGGCTTAGCCTGGCCCAACAGGAGTGGCTGCTCCGACCAAGTA 3597
      ::::::::::::::::::::
880 ArgThrIleAsnAlaArgGlnIyValIySArgIleSerSerIleuGln 896
      ::::::::::::::::::::
3598 CAAGACCAACAGCAGCTGAGAGATGAGAGAGCCAGGCTGGCCCGGG 3647
      ::::::::::::::::::::
896 uArgPheGlnIySArgIyGlnIyLeuValIleuAlaGlnIyIySArg 913
      ::::::::::::::::::::
3648 CCGGCAGAGGCTGAGAGAGAAATAGCCTCATCAGCAGCTTGGAGAGG 3697
      ::::::::::::::::::::
913 IyGlnIleuIleValIyIySArgIySLeuGlnIyIySerLeuAspProLeu 929
      ::::::::::::::::::::
3698 AGGTGCTCATCTGATATCGCAGAGCTCTGAGAAAGAGAGGGAG..... 3741
      ::::::::::::::::::::
930 SerIySArgIySArgIySLeuGlnIyIyThrAsnAlaLeuIySArgIy 946
      ::::::::::::::::::::
3742 AGCAAGAGATTGAAGCGGCTGTGATGCGCGAGTCAGAGAAAGCCANA 3791
      ::::::::::::::::::::
946 SLeuAspIleu..... 949
      ::::::::::::::::::::
3792 GCTGAGAGAGAGCTGCGCTCTGACAGCAGACAGCCAGACAGTGC 3841
      ::::::::::::::::::::
949 ..... 949
      ::::::::::::::::::::
3842 CAGAGCTCAGAACACAGCTCTCTGCGGAGAGAGTGCAGAGCTCC 3891
      ::::::::::::::::::::
950 .....GlnTyHisGlnIleuAlaGlnArg.....LysArgGln 960
      ::::::::::::::::::::
3892 GGGAGGAGAGCTGAGAAACAGCGGGTGGCTTCAGAGAACTTCCGAGCA 3941
      ::::::::::::::::::::
960 uPheGlnIyGlnIyLeuAspAlaLeuGlnIyArgLeuAsnMetIySArg 977
      ::::::::::::::::::::
3942 GCTGACCTCAGAGGCTGAGCTGCGAGAGAGCTGGCCAGAGATTGAAG 3991
      ::::::::::::::::::::
977 LTyIleuAspSer.....LysIySAsnGlnIyLeuIySArgIyGln 991
      ::::::::::::::::::::
3992 CGTGGCAGAGAGAGTTCTTCAGAAAGAGCAGCCCTTCCACCTGCAG 4041
      ::::::::::::::::::::
992 GlnArgHisVal..... 995
      ::::::::::::::::::::
4042 CTCGAGACACACAGCAGACAGCCCTGTGAGTGAAGCTGCTCCAGCTAA 4091
      ::::::::::::::::::::
996 ...LeuCyHisIleSerGlnIleuGlnSerCyMetAlaIySArgIyGln 1011
      ::::::::::::::::::::
4092 GCACCTCTCCAGCAGCTGACGCGCAGAGAGGCGCTGCCGAGAAAGGCC 4141
      ::::::::::::::::::::
1011 LeSerAlaGlnIleuAsnIySArgIyS..... 1019
      ::::::::::::::::::::
4142 ACCGTGAGAGAGCTGGAGAGCAGACAGCAGCCGCTGGGGGAGTGGCGGCA 4191
      ::::::::::::::::::::
1020 GlnLeuIleuGlnIyGlnIyGlnIyLeuIySArgAsnIleAspAspAsnIle 1036
      ::::::::::::::::::::
4192 GAGCTCTCTCGGGCCAGCGGGAGCTTGGGAGCTATTCCTCTGGCGCA 4241
      ::::::::::::::::::::
1036 uTyTyIyArgIySThrIyAlaAspValIyGlnIleuThrArgAspIleG 1053
      ::::::::::::::::::::
4242 GAAGGTGAGAGAGCAGAGCAGACAGCTCAGCAGCTGCGGCGCAGAACG 4291
      ::::::::::::::::::::
1053 IySerIyGlnIyArgIyLeuSerIleGlnIySerIleuSerAlaIleGln 1069
      ::::::::::::::::::::

```

```

TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and
TITLE OF INVENTION: Methods of Using Same
NUMBER OF SEQUENCES: 117
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe, Martens, Olson & Bear
STREET: 620 Newport Center Drive, Sixteenth Floor
CITY: Newport Beach
STATE: CA
COUNTRY: U.S.A.
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/056,200
FILING DATE: 30-Apr-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Fedirick, Michael F.
REGISTRATION NUMBER: 36,799
REFERENCE/DOCKET NUMBER: NIH054.001A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (714) 760-9502
TELEFAX: (714) 760-9504
INFORMATION FOR SEQ ID NO: 93:
SEQUENCE CHARACTERISTICS:
LENGTH: 9551 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1507..1644
FEATURE:
NAME/KEY: Introu
LOCATION: 1645..2511
FEATURE:
NAME/KEY: CDS
LOCATION: 2512..8070
US-08-056-200-93
alignment_scores:
Quality: 347.50 Length: 1138
Ratio: 0.616 Gaps: 43
Percent Similarity: 49.561 Percent Identity: 19.420
alignment_block:
US-09-538-396-2 x US-08-056-200-93 ..
Align seg 1/1 to: US-08-056-200-93 from: 1 to: 9551
73 gylgltThrlGlnThrlysglgnllleYslLeuArGPheIysTrAlaAl 89
||||| ::::~: ||| :|||:~: ::
2824 GGAGAGAGACCAAGTGTGAACAAGACCGACTTGAAACAGCGCACGAGA 2873
89 agLyAsApVaIValCySlleArGserPheGlnLeuThrGlnYsAlas 106
::::~: ||| |::~: |||
2874 GCCGCCGCAGAGAGAGCTGTGGCGCAAGCAAGAAATGGCMAGAA ... 2919
106 erYsmeTGIunPhelysAlalIeqluSerVallengInTrIlleasnPro 122
::::~: ||| :|||:~: ||| |::~: |||
2920 . CGGAGAGAGCCCGCGCAGAGAGAGAGACGCTGCAGAGATTCGCAAAGT 2967
123 HlsThrlgylulYsValCySleuSerTyrrGcySAlasPmetaaP.. 138
||||~: ||| :|||:~: |||
2968 CACGAAACTGTAGAGAGTTT.....CCAGCAGAAAGAGA 2999

```


4335 GAGCGAGAGAGAGAGAGAGCGCCACGAGCACTGAGCGCGAGAGAGC 4384
743 eprylsleuargthrlletryasplalalyvalylsleuvalglugluthr 759
4385 AGGAAAGCGC.....GAGCAGCGCGCTGAAAGCGCGAGAGAGAG... 4422
760 lleroleuualagluysasnleuasnlnhisleuualasplusergl 776
4423GAGAGAGCGCTGACAGCGCGCTGAAAGCGCGAGCATGA 4460
776 nlysalaglnalaphaeasplleuglyvalleuualhisvalglm 793
4461 GGAAGAGAGCGCGAGAGAGAGCA..... 4485
793 etasplaryasplalalyvalleuualleuuglnprothrapsplthrlleasp 809
4486GCTGAGAGAGAGAGAGAGAGCGCGCGCGAGCTTAAAG 4524
810 Arghlsvalnhsigluegln..Glnleuvalylsugluvalglu.... 823
4525 AGCGCGATCCGAGAGTGGCAGTGGCAGCTAGAAAGCGAGCGAGAGAGCAG 4574
824 AspleuugluthrAlaleuaspserSerGlyArgGlyValylsleu 840
4575 GCAAGAGCAAACTCTTACTCGAGCGCCCGCAAGAGAGAGCGAGAGCGGC 4624
840 lueglueglnleuasnphleuuglnargthrargasphrleu 856
4625 CGCAGAGAGAGAGAGCA.....AAGAGCGCGCGAGAGAGTGA 4662
857 lIlevalgluvalasplleuargasplnhsargmetleuasnleu 873
4663 CTGCAATGGCAGAGAGAGAGAGAGCG...GCTCAGCGCGAGAGAGAGAGA 4709
873 pmetsereralaglnvalargthrphisasnAlahrglugluthrval 890
4710 GAGAGAGCGCGCGAGCTTCACATGGCAGTGGCAGCGAGAGAGAGAGCG 4759
890 ysAlaserSerlleuugluarphnegl.....LysSer 901
4760 AGAGGCGCGCT.....CAGAGCGCTGCGCGAGCGCGCGCGAGAG 4803
902 Glugluglueuvalleuualaglugluthrlysgluglueuvalgl 918
4804 CAGCGGAGAGAGAGAGAGAGCGCGAGAGAGAGAGAGAGAGAGAGAGAG 4842
918 ulslyslleuuglueuargluserleuasprouserlysgluthrlysg 935
4843CGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4882
935 eRleuuglueuugluthrAsnAlaleuulysgluleu..... 947
4883 AGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4932
948 asplugluthrHisgluleuualagluarGlyArgGluPhneglngl 964
4933 GAAG 4982
964 uleuaspalaleu.....GlyArgL 971
4983 GAG 5032
971 euasnmetlysllelysglytyrleuaspserlyslasngluthrlysl 987
5033 GCGAGCAAAAATGAGGTGCAACTAGAAAGAGAGAGAGAGAGAGAGAGAG 5079
988 LysgluleuugluthrArgHisvalleu..CysHisserGlnleuuglnse 1003
5080CGCCACAGCGCTGTACGCCAGAGAGAGAGAGAGAGAGAGAG 5114
1003 rCysmetAlalyglnlnarglleseralagluLeuasnlyserlysg 1020
5115 GCAGCTGAG 5158

1020 luleuuglueglnlyglnleuLeuulysargasnIleasplasnleu 1036
5159 AGCTACAG 5208
1037 LysThrargluthrlyslAlasplvalgluglueuulthrargaspllegl 1053
5209 CAATACCGC.....GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5241
1053 userleuugluthrleuuserlleuglyserleuserAlatlegl 1070
5242GAG 5263
1070 lAspleuulysargHisserGlnleuulysgluargleuasnserGluPh 1086
5264 AACGAG 5298
1087 AsnargthrgluthrleuuserValtyrglnserAsnIleSerlysnI 1103
5299CAATA 5303
1103 slysglnleuulysleuserGlnlyslaspllegluthrlyslarglytt 1120
5304 TCGAGAGAGATAGAGAGTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5337
1120 hrasnlnphleuuglnleuulysrthrtrglumetAlasnllyaspleu 1136
5338CAGCTGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5382
1137 aspargthrytyrthrAlaleuaspIlysalaleuMetargphneIserme 1153
5383 GAG 5432
1153 tlysmetglugluleuasnlysllellelysglueuulthrtrglnlnthr 1170
5433 GCTGAG 5482
1170 ytrgglglnasp 1174
5483 ACCGCAAAAAGAGC 5496

Thu Nov 8 09:02:35 2001

us-09-538-396-2.rn1

1207 MetArgLYArgCysSerAlaGlyGlnValLeuAlaSerLeu11 1223
 152 ATGCGAGGGAGGTGTAGTGTGGCCAAAGGTCTTGCATCATCATCAT 201
 1223 GATGLeuAlaLeuAlaGlnThrPheCysLeuAsnCysGlyIleLeuAla 1240
 202 ACCATTGGGGTGGTGGTGAACCTTTTGTCTCAACTCGGGAATTCAGCAC 251
 1240 euAspGluProThrThrAsnLeuAspGlyProAsnAlaGluSerLeuAla 1256
 252 TGGATTAACCTCCACAACTAGACGGCCCAACAGTGAAGTCTTGGCT 301
 1257 AlaAlaLeuLeuArgIleMetGluAlaArgLYGlyGlnGluAsnPhel 1273
 302 GCGAGCTGTAGATTAATGAGGACAGAAAAGGCCAAGAAATTTCA 351
 1273 nleuIleValIleThrHisAspGluArgPheAlaHisLeuIleGlyGln 1290
 352 ACTAATTCATTCACACAGATGAGGATTTGCTCAATATATGGCCAAC 401
 1290 rglInleuAlaGlyGlyTyrTyrArgValSerLYsAspGluAsnGlnHis 1306
 402 GCCAGATGCGAAAAGTATATCGATCTCAAGACAGCACATCAGCAC 451
 1307 SerIleIleGluSerGlnGluIlePheAsp 1316
 452 AGTATATTCACAGCTCAAGAAATTTGAT 481
 seq_name: gb_est30:AU100679

seq_documentation_block:
 LOCUS AU100679 706 bp mRNA EST 22-AUG-2000
 DEFINITION AU100679 Rice callus Oryza sativa subsp. Japonica cDNA clone C11847
 ACCESSION AU100679
 VERSION AU100679.1 GI:9864929
 KEYWORDS EST.
 SOURCE Oryza sativa subsp. Japonica.
 ORGANISM Oryza sativa subsp. Japonica.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Euphorbiaceae; Oryzaceae; Oryza.
 REFERENCE 1 (bases 1 to 706)
 AUTHORS Sasaki, T. and Yamamoto, K.
 TITLE Rice cDNA from callus (2000)
 JOURNAL Unpublished (2000)
 COMMENT Contact: Takuji Sasaki
 National Institute of Agrobiological Resources
 Rice Genome Research Program
 2-1-2 Kannondai, Tsukuba
 Ibaraki,
 Japan 305
 Tel: 0298-38-7441
 Fax: 0298-38-7468
 Email: tsasaki@abr.affrc.go.jp
 PROJECT -"RGP"
 C11847_6Z.
 FEATURES
 source Location/Qualifiers
 1..706
 /organism="Oryza sativa subsp. Japonica"
 /strain="cultivar Nipponbare, sub-species Japonica"
 /db_xref="taxon:39947"
 /clone="C11847"
 /clone_1lb="Rice callus"
 /note="Vector: pBluescript II SK+; Site_1: SalI; Site_2:
 NotI; cDNA prepared from rice callus mRNAs by using
 oligo(dt) as a primer and ligating to the SalI-NotI site
 of pBluescript II SK+ phagemid."
 BASE COUNT 262 a 114 c 175 g 152 t 3 others
 ORIGIN

alignment_scores:
 262 a 114 c 175 g 152 t 3 others

Quality: 700.00 Length: 233
 Ratio: 3.500 Gaps: 4
 Percent Similarity: 85.837 Percent Identity: 64.378

alignment_block:
 US-09-538-396-2 x AU100679 ..

Align seg 1/1 to: AU100679 from: 1 to: 706

781 PheAspAspLeuGluGlyValLeuAlaHisValGlnMetAspArgAsp 797
 17 TTGATGATTTTGTCACTGCTNCTTCCCTCAAGTTAAATGCGCAAGATG 66
 797 laval.GluAlaLeuLeuGlnProThr.AspThrIleAspArgHisVal 813
 67 CAGTGGCAGGTATGCTACAGCCAGTGAACGATTAAGCATGATGAGCAGTNC 116
 813 lsgIleGlnGlnLeuValLYsGluValGluAspLeuGlyTyrAlaLeu 829
 117 AAGAAATCCAGCAGTTAGCACCCNCAAGTTGAAATCTTGATATTAAGCTT 166
 830 AspSerSerGlyArgGlyValLYsSerLeuGlnGluIleGlnLeuGlu 846
 167 GATGTTGTCGCGCAAGGTGTTAATCTTGGACCAATCCAGTTGGAGTT 216
 846 uAsnPhelGluArgThrArgAspThrLeuIleValGluValAspAsp 863
 217 GAACCTGTGTCAGAGACAGCAAGGACACATTAACATGAAGTGGATGATC 266
 863 euArgAspGlnHisArgMetLeuAsnGluAspSerSerAlaGlnVal 879
 267 TAAGGATCAACAAGAACACTTACCGATGCTGCTGACTAATGCTCAGATG 316
 880 ArgTyrHisAsnAlaArgGlnGluLYsValLYsAlaSerSerIleLeuG 896
 317 CGATGCGATGATATTAGGGAAGAGAGAACTAAAGCTTCAGTCCAGTCA 366
 896 uArgPheGlnLYsSerGlnGluGluValLeuAlaGlnGluLYsG 913
 367 CAATATTCAAAAAGCTGAGAGAGATTTGGGACATCTAGCTGAGAGAGAAG 416
 913 lglInleuIleValGluLYsLYsLeuLeuGlnGluSerLeuAspProLeu 929
 417 AAAAAGTAACTATAGAAAGAACATTTAGAGAGATCTTGGTGGCTGTG 466
 930 SerLYsGlnLYsSerLeuLeuGlnGlnTyrAsnAlaLeuLYsGlnLY 946
 467 TCAAAAGAGAGAGAAAGCTTTGCAAGACATGAAGCTTTGAAGGAAA 516
 946 sleuAspGluGlnTyrHisGlnLeuAlaGluArgLYsArgGluPheGln 963
 517 GCTTGATCAGAGATATCATCACTGCGCAGAAAAGAAAAGGAATTCACGC 566
 963 lngluLeuAspAlaLeuGlyArgLeuAsnMetLYsIleLYsGlyTyrLeu 979
 567 AGGAATTTGATGCATCGACACACACAAATAGCGTATCAAAAGGTACCTG 616
 980 AspSerLYsLYsAsnGluLYsLeuLYsGlnLeuGlnGlyArgHisVal 996
 617 AATTCAAAAGAGGCGGAGCTTAATGAACACTACAGGAAAAA..... 658
 996 uCysHisSerGlnLeuGlnSerCysMetAlaLYsGlnGlnArg 1010
 659 ...CATACTCACTTCAGTCTGATCTACAAAAA..... 697
 seq_name: gb_est77:BE637160
 seq_documentation_block:
 LOCUS BE637160 609 bp mRNA EST 25-AUG-2000
 DEFINITION WHE1806_H09_O18Z Secale cereale anther cDNA library Secale cereale
 cDNA clone WHE1806_H09_O18, mRNA sequence.
 ACCESSION BE637160
 VERSION BE637160.1 GI:9920271

KEYWORDS EST.
SOURCE rye.
ORGANISM Secale cereale
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
; Triticeae; Secale.
1 (bases 1 to 609)
REFERENCE
AUTHORS Anderson, O.D., Butler, E., Chao, S., Choi, D.W., Close, T.J., Fenton
, R.D., Gustafson, J.P., Han, P.S., Hsia, C.C., Kang, Y., Lazo, G.R.,
Miller, R., Rausch, C.J., Ross, K., Seaton, C.L. and Tong, J.C.
The structure and function of the expressed portion of the wheat
genomes - Another cDNA library from rye
JOURNAL
COMMENT Unpublished (2000)
Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oanderson@pv.usda.gov
Sequence have been trimmed to remove vector sequence and low
quality sequence with phred score less than 20
Seq primer: Stragene SK primer.
Location/Qualifiers
1..609
/organism="Secale cereale"
/cultivar="Blanco"
/db_xref="taxon:4550"
/clone="WHE1806_H09_018"
/clone_lib="Secale cereale anther cDNA library"
/tissue_type="Anther"
/dev_stage="Adult plant before anthesis"
/lab_host="E. coli SOLR"
/note="Vector: Lambda Uni-ZAP XR, excised phagemid;
Site 1: EcoRI; Site 2: XhoI; Plants were grown in the
greenhouse. Anthers were harvested and pooled from early
meiosis to late meiosis. The tissue, total RNA, and
poly(A) RNA were prepared (Butler, Ross and Gustafson) at
University of Missouri, Columbia. A cDNA library was
made, and the cDNA clones were in vivo excised to give
phagemid phagemids in the TV Close lab (Choi, Close,
Fenton) at the University of California, Riverside.
Plasmid DNA preparations and DNA sequencing were performed
in the OD Anderson Lab (all other authors)."
BASE COUNT 219 a 100 c 148 g 142 t
ORIGIN
alignment_scores:
Quality: 689.00 Length: 200
Ratio: 3.786 Gaps: 0
Percent Similarity: 91.000 Percent Identity: 68.500
alignment_block:
US-09-538-396-2 x BE637160 ..
Align seg 1/1 to: BE637160 from: 1 to: 609
872 GUAspmerserserlagnvalargtphhsasnaalarglglulyl 888
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
8 GAGGATCTATCAATGCTCAGATGCGATGCGATCTTGGAGAGAAA 57
888 svaLysAlaserserlreulgluarphgncInLyssergluglul 905
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
58 ACRAAAGCCTCAAGTATGCTGAAGTCAAAAGCGCGAAGAGATT 107
905 euValleuenaalaglulysglulnleullevalglulyslvsl 921
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
108 TGGTACATTTGCCGAGAGAAAGAGCAACTGATCCAGATCAGAACAT 157
922 leuglulglulserleuasprtolserlsglulysglulserleu 938
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
158 TTGAAGAGAGCTCTGTTCCATTGTCAAAGAGAGAGAAAGCTTGTTCA 207

938 ngLurYrsnaLaleuYsGlnLysleuaspeluglulurYhsglnleu 955
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
208 AGAGTATAAGCTTTGAAGAAAGGTTGATCAGAGATATGATCAGCG 257
955 lagluarLysargluphnglnglngluLeuaspAlaleuLysar 971
258 CAGAAAGAAAAGGGGATTCAGCAAGAAATCGATGATCTTGAAACCTT 307
972 AanmetLysleLysglYlYrleuasprserLysAsnnglulYleu 988
308 AACACAGCATCAAAAGGGTATCTGATTCMAAAGAAAGTAGAACCTTAA 357
988 sglulenglnglYarGhlsvalleucYhssercInleuInserCysm 1005
358 TGAATTCAGGAAGCAATACCTTATCCCTATCTCAGTTACGAATG 407
1005 eAlalysglnlnglYarGlseserlagnleuaspLysserlglul 1021
408 AGCAAGAAAGCAAGACATCGCGGTTGACCTTGACAGCAAGCAAGCTA 457
1022 leuglnglYlnglYlnleuYsarYasnleaspaAsnleuYery 1038
458 TTACGAGCAAGATCAATTAAGAAATATGATGACAACTGAACCTA 507
1038 rArgLysThrLysAlaaspaIglulnleuthrArgasplleglulserL 1055
508 CAGGAAACAAAGGCTGAAGTGATGATCTACTGATGATCGAGTTAC 557
1055 englulnLarglleuaserlleglYserleuSerAlaleuLaa 1071
558 TTCAATATATGCTTCATCGCAGATGCTATAGAGCTGAT 607
seq_name: gb_est74:BE449712
seq_documentation_block:
LOCUS BE449712 486 bp mRNA EST 26-JUL-2000
DEFINITION BE449712 tomato root, plants pre-anthesis, Cornell University
Lycopersicon esculentum cDNA clone CLEVI1K13 similar to
strong similarity to A. thaliana putative RAD50 DNA repair protein,
mRNA sequence.
ACCESSION BE449712
VERSION BE449712.1 GI:9455215
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
1 (bases 1 to 486)
REFERENCE
AUTHORS van der Hoeven, R.S., Garvin, D., Matern, A.L., Holt, I.E., Liang, F.,
Upton, J., Hansen, T., Craven, W.B., Bowman, C.L., Alm, S., Konning, C.M.,
Fraser, C.M., Martin, G.B., Giovannoli, J.J. and Tanksey, S.D.
Generation of ESTs from tomato root tissue
Unpublished (1999)
JOURNAL
COMMENT Contact: David Frisch
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfrisch@CLEMSON.EDU
5 prime sequence.
FEATURES
location/Qualifiers
1..486
/organism="Lycopersicon esculentum"
/cultivar="FA496"
/db_xref="taxon:4081"
/clone="CLEVI1K13"
/clone_lib="tomato root, plants pre-anthesis, Cornell
University"
/tissue_type="root"

/dev_stage="plants in pre-anthesis stage"
 /note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
 XhoI; supplier: Tanksley; tissue supplied by Dave Garvin
 (USDA-ARS, Ithaca, NY 14850)."
 ORIGIN
 BASE COUNT 152 a 99 c 125 g 110 t

alignment_scores:
 Quality: 655.00 Length: 134
 Ratio: 4.962 Gaps: 0
 Percent Similarity: 98.507 Percent Identity: 90.299

alignment_block:
 US-09-538-396-2 x BE449712 ..

Align seg 1/1 to: BE449712 from: 1 to: 486

```

1 MetSerThrValAspLysMetLeuIleLysGlyLeuArgSerPheAspPr 17
|||||
84 ATGACCTGTAGATAGATGCTCTAAGAGGCGATCCGAGATTGACCC 133
|||||
17 CAspAsnLysAsnValIleThrPhePheLysProLeuThrLeuIleValg 34
|||||
134 GGAGAGCAAAAAAGCTTATCACCCTTTCCGACCATTAATCTCATCTCG 183
|||||
34 LYPProAsnGlyAlaGlyLysThrIleIleGluCysLeuLysLeuSer 50
|||||
184 GAGCTACAGCGCGCCGAAAAACTACTGTAATTGAGTCTTGAAAGTTGCT 233
|||||
51 CysThrGlyLysLeuProAsnSerArgSerGlyHisThrPheValH 67
|||||
234 TGCACTGGTGTATGTCGCCCAATGTCCTGCTGCTACGTTATCTCA 283
|||||
67 AspProLysValAlaGlyLysThrGluThrLysGlyGlnIleLysLeu 84
|||||
284 CGACCTTAGAGTGCTGAGAGACGAGACAAAGGCGCAGATTAAGCTGC 333
|||||
84 TGPheLysThrAlaAlaGlyLysAspValValCysIleArgSerPheGln 100
|||||
334 GGTTCAAAACAGCAGCAGGAGAAAGATGTGTGCTATGAGCTTTTTCAG 383
|||||
101 LeuThrGlnLysAlaSerLysMetGluPheLysAlaIleGluSerValle 117
|||||
384 CTAACTCGAAGGCGCAACAAGATGGAATACAGCCATTGAGAGGTTCT 433
|||||
117 uGlnThrIleAsnProHisThrGlyLysValCysLeuSerTyrArgC 134
|||||
434 ACAAGCTTAATATCTCTACACGCGAGAGAAAGTTGCTTAAGTTATGAT 483
|||||
134 YS 134
||
484 GT 485

```

seq_name: gb_est75:BE515984

seq_documentation_block:
 LOCUS BE515984 634 bp mRNA EST 08-AUG-2000
 DEFINITION WHE0607_E01_J012A wheat ABA-treated embryo cDNA library Triticum
 aestivum cDNA clone WHE0607_E01_J01, mRNA sequence.
 ACCESSION BE515984
 VERSION BE515984.1 GI:9740010
 KEYWORDS EST.
 SOURCE bread wheat.
 ORGANISM Triticum aestivum
 Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
 ; Triticeae; Triticum.
 1 (bases 1 to 634)
 REFERENCE
 ANDERSON, O.D., CHAO, S., HAN, P.S., HSLA, C.C., JOHNSON, R.R., KANG, Y.,
 LAZO, G.R., MILLER, R., RAUSCH, C.J., SEATON, C.L., TONG, J.C., VERHEY,
 S.D., and WALKER-SIMMONS, M.K.
 The structure and function of the expressed portion of the wheat

JOURNAL
 Unpublished (2000)
 COMMENT
 Contact: Olin Anderson
 US Department of Agriculture, Agriculture Research Service, Pacific
 West Area, Western Regional Research Center
 800 Buchanan Street, Albany, CA 94710, USA
 Tel: 5105595773
 Fax: 5105595818
 Email: oanderson@pw.usda.gov

Sequence have been trimmed to remove vector sequence and low
 quality sequence with phred score less than 20
 Seq primer: Clontech Matchmaker 3' AD primer.
 Location/Qualifiers

FEATURES

source
 1. 634
 /organism="Triticum aestivum"
 /cultivar="Brevor (soft, white, winter, common wheat)"
 /db_xref="taxon:4565"
 /clone="WHE0607_E01_J01"
 /clone_1bp="Wheat ABA-treated embryo cDNA library"
 /tissue_type="Seed embryo"
 /dev_stage="Mature dormant seeds"
 /lab_host="E. coli DH12S"
 /note="Vector: pGAD10; Site 1: EcoRI; Site 2: XhoI;
 Embryos were cut from mature, dormant seeds and imbedded in
 25 microm ABA (abscisic acid) in 5 mM Mes buffer, pH 5.7,
 for 12 hr at 22 C. The tissue, total RNA, and poly(A)
 RNA were prepared by Steven Verhey in M.K.
 Walker-Simmons's lab (USDA-ARS, Washington State Univ.,
 Pullman, Washington 99164-6420). A cDNA library was made
 by Clontech using a combination of random and oligo dT
 primers. Library was plated and archived by Russell
 Johnson (Colby College, ME/Walker-Simmons' lab). Plasmid
 DNA preparations and DNA sequencing were performed in the
 OD Anderson lab (all other authors)."

BASE COUNT 146 a 144 c 111 g 233 t
 ORIGIN

alignment_scores:
 Quality: 630.00 Length: 203
 Ratio: 3.600 Gaps: 0
 Percent Similarity: 86.207 Percent Identity: 59.113

alignment_block:
 US-09-538-396-2 x BE515984/rev ..

Align seg 1/1 to reverse of: BE515984 from: 1 to: 634

```

502 AspTyrAspSerIleIleSerGlnLysArgThrGluValTyrSerLeuGln 518
|||||
632 GACTATGATTGATTATATAGTACAGGCGCTCAGAGATATATACCTTGA 583
|||||
518 uGlnLysIleLysValLeuLeuArgGluLysAspIleIleAsnArgAsnA 535
|||||
582 TCACAAGATTAAGAAGCATTCACAGAGAAAGATTAACATGACGACTGANG 533
|||||
535 LAAspGluArgValLysLeuGlyLeuLysLysAspAlaLeuLysSer 551
|||||
532 CTGATGACAGAGAGTAAATTAAGCAAGAGGATGAGTGAAGATGTC 483
|||||
552 LysAspLysLeuAsnGluIleValAsnGluHisLysAspLysIleLysLys 568
|||||
482 AAGAAGAAAGCTTAAGAAGATATATGATGACATTAAGGATTAATTTAGAAG 433
|||||
568 sValLeuArgGlyArgAsnProPheGluLysAspMetLysLysGluIleA 585
|||||
432 TGTCTTAAGGAGAGCGTCTCATGAGAGAGATGTCAAGAAGAGATTA 383
|||||
585 snglnAlaPheThrProValAspLysGluTyrAsnGluLeuArgSerLys 601
|||||
382 CTCAGGCTTTCGGTCTGCTGACATGCAATACAAATGATTGACATCAAAA 333
|||||
602 SerGlnGluAlaGluGluLeuLysPheThrGlnSerLysValThrAs 618

```

```

|||||
332 TCTCAGGAGCGGAGAACACTTGTAATTTGGCACAATAATGATGATCC 283
618 PALAARGLUGInLeuThrLysLeuArgArgAspMetAspAlaLysArg 635
|||||
282 TGCTAAAGCCACTTGGCAAGCTCCAAAAGTTTGGATGCAAAAAGAA 233
635 rghLeuAspSerLysLeuInserLleLeuGlnLleSerAlaAsnVal 651
|||||
232 AGCATCTGAACCTGAACTCATTCATGCTAGTATCTGTTACATC 183
652 AspMetPheProLysValLeuGlnAspAlaMetAsnLysArgAspGlu 668
|||||
182 AACGCTTACCCCAAGATTCTGAAAGATGCCATGCGACGAGAGATTAACA 133
668 nLysArgLeuGlnAsnPheAlaAsnGlyMetArgGluMetLeuAlaPro 685
|||||
132 GACAAATATTTCACTATGCTAAGGAAATGCGCAATGTATGAACCT 83
685 heGlnHisLeuAlaArgLysAsnHisValCysProCysCysGluArgAla 701
|||||
82 TTGAAAGAGTGCGCCGACAGCATCAAGTGTCTGTGTGATGCTGCT 33
702 PheThrPro 704
|||||
32 TTCACACCT 24
seq_name: gb_est30:AU122773

```

```

seq_documentation_block:
LOCUS AU122773 901 bp mRNA EST 23-Oct-2000
DEFINITION AU122773 MAMMAL Homo sapiens cDNA clone MAMMA1003095 5', mRNA
sequence.
ACCESSION AU122773
VERSION AU122773.1 GI:10947489
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 901)
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isozaki,T.
JOURNAL HRI human cDNA project
COMMENT Unpublished (2000)
Contact: Takeo Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix
Research Institute; cDNA library construction; Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
FEATURES
Source
1. 901
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MAMMA1003095"
/clone_lib="MAMMA1"
/tissue_type="mammary gland"
/note="Vector: pME18SFL3"
BASE COUNT 326 a 145 c 198 g 229 t 3 others
ORIGIN

```

```

alignment_scores:
Quality: 627.50 Length: 304
Ratio: 2.716 Gaps: 8
Percent Similarity: 75.987 Percent Identity: 44.408

```

```

alignment_block:
us-09-538-396-2 x AU122773
Align: seg 1/1 to: AU122773 from: 1 to: 901

```

```

9 lleLysGlyIleArgSerPhe.....AspProAspAsnLysAsnValIle 23
|||||
5 ATTGTGGCGGTGGGAGTTTGGAAATAGAGCAAAAGATTAAGCAAAATAT 54
23 eThrPhePheLysProLeuThrLeuIleValGlyProAsnLysIleGlyL 40
55 CACTTTCTTCAGCCCTTACAAATTTGTTGGAGCCCAATGGCGGGA 104
40 ySThrThrIleIleGluCysLeuLysLeuSerCysThrGlyGluLeuPro 56
|||||
105 AACAGCAACATCTTGAATGCTAAATATATTGTTACAGGAGATTTCCT 154
57 ProAsnSerArgSerGlyHisThrPheValHisAspProLysValAlaG 73
|||||
155 CCTGGAACCAAA...GGAATACATTTGTACAGCATCCCAAGTTGTCTCA 201
73 yGluThrGluThrLysGlyGlnIleLysLeuArgPheLysThrAlaIle 90
202 AGAAACAGATGTGAGAGCCCAAGATTCGTGCAATTTGTCGATGCAATG 251
90 LysAspAlaValAlaCysIleArgSerPheGlnLeuThrGlnLysAlaSer 106
|||||
252 GAGAACTTATGCTGTGCAAAAGATCTATGCTGTGTACTCAGAAAGCAAA 301
107 LysMetGluPheLysAlaIleGluSerValLeuGlnThrIleAsnProH 123
|||||
302 AAGACAGAAATTTAAACTGTGAAGAGCAT...ACTAGAACAAAGCA 348
123 sThrGlyGluLysValCysLeuSerTyArgCysAlaAspMetAspArg 140
349 T...GGTGAAGAGTCAGTCAGTCGTGAGTGTGAGTAATGTGACCGAG 395
140 IuLeProAlaLeuMetGlyValSerIleValIleValLeuGlnLysAla 156
|||||
396 AATGATCAGTCTCTTGGGGTTCCAGAGCTGTGCTAAATATATGTCATT 445
157 PheValHisGlnAspGluSerAsnTrpProLeuGlnAspProSerThr 173
|||||
446 TTCTGTCATCAGAGAAATCTTAATTCCTTAAGTGAAGAAAGGCTT 495
173 uLysLysLysPheAspAspIlePheSerAlaThrArgTyThrLysAla 190
496 GAAGCAAAAGTTTGATGAGATTTTTCAGCAACAAGATACATTAAGCCT 545
190 euGluValIleLysLysLeuHisLysAspGlnMetClnLysIleThr 206
|||||
546 TAGAAACACTTCGCGAGGTACGTACAGACAAAGTGAAGTAAGTAAGAA 595
207 PheArgLeuLysLeuGlnAsnLeuGlnThrValLysAspGlnAlaHis 223
596 TATCAATGGAATTAATATCTGAAGCAATATTAAGGAAAAAGCTGTGA 645
223 sLeuArgLysAsnIleAlaGlnAspGlnLysSerAspAlaSerLys 240
646 GATTCGATCAGATTAACAAGTAAGAACGCCAGTTAAACATCTTCAAG 694
240 eGlnMetGluGlnLeuLysGlyLysIleCysGlyThrGluArgGlu 256
695 .....GAAAT 700
257 LeuGln...MetGluThrSerLeuAspGluLeuArgArgLeuGlnLys 272
701 GTCAAAATCCTATAGATGAAATGATTCATTAAGAAATGCTATAAG 750
272 nIleAspIleLysAlaThrGluArgSerThrLeuLeuThrGlnGlnHis 289
751 AATGACATATAT.....CTCTCTAAATAAATGAAGAACTGTGACAATG 790

```

```

289 JulysaenulalaleuserglugluangluspThrasplunguleu 305
|| ..... |||||||.....:|||||
791 M...ATTMAAGCTTGATGACCCGAAGAACAATGAGAAAGATATA 837

306 Metjutrpgln 309
::::::::::
838 GTGAAGTGGAG 849

seq_name: gb_est69:BE013657

seq_documentation_block:
LOCUS BE013657 531 bp mRNA EST 09-JUL-2000
DEFINITION 125081 MARC_1Pig Sus scrofa cDNA 5', mRNA sequence.
ACCESSION BE013657
VERSION BE013657.1 GI:8274637
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryote; Metazoa; Chordata; Granata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Cetartiodactyla; Suidae; Slidae; Sus.
REFERENCE
1 (bases 1 to 531)
Fahnenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
and Keel,J.W.
Design and use of two pooled tissue normalized cDNA libraries for
EST discovery in swine
Unpublished (2000)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@easill.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.98090.e. Vector identified by cross_match with the -minscore 18
and -mismatch 12 options.
PCR Primers
FORWARD: AGGAACGACTATGACCAT
BACKWARD: GTTTCCAGTCACGCAGC
Plate: 47 row: J column: 19
Seq primer: ATTTAGGTGACCATATG.
Location/Qualifiers
1..531
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_id="MARC_1PIg"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
library made from pooled tissue from day 11, 13, 15, 20,
and 30 embryos"
BASE COUNT 174 a 98 c 121 g 138 t
ORIGIN
alignment_scores:
Quality: 599.50 Length: 176
Ratio: 3.893 Gaps: 1
Percent Similarity: 87.500 Percent Identity: 66.477

alignment_block:
US-09-538-396-2 x BE013657 ..

Align seg 1/1 to: BE013657 from: 1 to: 531

1117 LysatgytYThAsngInPhLeuIngInLysrThrHglUMetAlaAs 1133
::::::::::::::::::::::::::::::::::::::::::::::::::
2 GAAGAGTAAGCAATGCGTGAATGTTATAGAACACACAGACTTGTGAA 51

1133 nlysaapleuAsparqTYrTYrThAlaLeuSpLySaLaLeuMetArp 1150
|||||
52 CAAGGATCGACATTTATTATTAAGACCTTACCAAGCGATTAATGAAT 101

```

```

1150 heh1sermetlysmetglugluileasnyllelleysgluleutrp 1166
|||||
102 ttccacg1atgaaatgaaagaaatacaaaattttctgtgcctctgg 151
1167 g1ng1nthrttyrarg1ylg1nasp1leaaptyr1leser1leasnseras 1183
|||||
152 cgaagactatnctgtgcacagatattgaaatcatnagaattcgactga 201
|||||
1183 p.....Serg1ug1y1a1a1y1thra1gser1tyr1ser1tyra 1195
|||||
202 tcccgatgaaatg1atctcagcttctgataagacggcgaattcaactacc 251
|||||
1195 rgvalyalmetg1nthrtg1yasp1alag1luenc1luwetr1rg1yarg1ys 1211
|||||
252 ggg1tgatg1atgctaaaggagacacgcttgatn1yrg1gagacggctgc 301
|||||
1212 ser1ag1g1n1lysy1al1eua1aser1leu1le1learg1eua1aleua1 1228
|||||
302 ag1gc1g1gacag1ag1st1at1agctt1ac1cat1ct1gcgct1agcct1ggc 351
|||||
1228 ag1uthr1phe1cys1leua1ncys1g1yl1leu1alal1eua1sp1u1p1ro1th1t 1245
|||||
352 tgaacat1ctgct1aa1c1tgc1at1tgc1ct1agcct1g1gag1cc1a1c1a 401
|||||
1245 hr1asn1eua1sp1g1y1p1ro1asn1ag1u1ser1leua1al1al1aleu1leua1n1g 1261
|||||
402 caaa1tt1t1ac1c1c1tga1aa1at1tga1n1t1ct1tgc1ac1at1gct1g1tg1ag 451
|||||
1262 11em1et1g1ua1ar1p1lys1g1yng1n1uas1n1phe1gn1leu1lle1val1leth 1278
|||||
452 at1aat1p1aaa1ag1tgc1t1cac1ag1c1g1t1a1c1t1tgc1a1c1t1tgc1t1a1t1y1ac 501
|||||
1278 r1h1s1asp1g1u1ar1g1phe1al1h1s1leu1le 1287
|||||
502 tcatg1atg1aa1gat1tt1gtg1aa1c1t1c1t1a 529
|||||
seq_name: gb_est41:AW032232

seq_documentation_block:
LOCUS AW032232 588 bp mRNA EST 15-SEP-1999
DEFINITION EST2735686 tomato callus, TAMU Lycopersicon esculentum cDNA clone
c1ec35n16, mRNA sequence.
ACCESSION AW032232
VERSION AW032232.1 GI:5890988
KEYWORDS EST.
SOURCE
ORGANISM tomato.
Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
1 (bases 1 to 588)
Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E.,
Liang,F., Upton,J., Craven,M.B., Bowman,C.L., Ahn,S., Roming
,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
Generation of ESTs from tomato callus tissue
Unpublished (1999)
Contact: David Frisch
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfrisch@clemson.edu
5 prime sequence.
Location/Qualifiers
1..588
/organism="Lycopersicon esculentum"
/cultivar="TA96"
/db_xref="taxon:4081"
/clone="c1ec35n16"
/clone_1lb="tomato callus, TAMU"

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2001, 12:58:30 ; Search time 5823.37 Seconds
(Without alignments)
11931.430 Million cell updates/sec

Title: US-09-538-396-1

Perfect score: 4492

Sequence: 1 aattcgacacgagtgcattcc.....atctgacatcaaaaaaaaa 4492

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1344157 seqs, 7733874588 residues

Word size : 30

Total number of hits satisfying chosen parameters: 1

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: gb_ba1:*
17: gb_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_rod:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_v1:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_v11:*
59: gb_v12:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_r01:*
95: gb_r02:*
96: gb_in4:*
97: gb_pr10:*
98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4492	100.0	4492	9 AX047235	AX047235 Sequence

ALIGNMENTS

RESULT 1
AX047235

LOCUS	AX047235	4492 bp	DNA	PAT	15-DEC-2000
DEFINITION	Sequence 1 from Patent W00068404.				
ACCESSION	AX047235				
VERSION	AX047235.1	GI:11876516			
SOURCE					
ORGANISM	Zea mays.				
REFERENCE	Eukaryote: Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoidae; Andropogonae; Zea.				
AUTHORS	1 (bases to 4492)				
TITLE	Mahajan, P. B. and Shl. J.				
JOURNAL	Matz, Rad50 orthologue and uses of thereof Patent: WO 0068404-A 1 16-NOV-2000; PIONEER HI-BRED INTERNATIONAL, INC. (US)				
FEATURES					
source	1..4492				
	/organism="Zea mays"				
	/db_xref="taxon:4577"				
CDS	292..4242				
	/note="unnamed protein product"				
	/codon_start=1				
	/protein_id="CAC18962.1"				
	/db_xref="GI:11876517"				
	/translation="MSTVDKMLKIGISFDPDNKNVITFEPLLLTIGVPGACGKTTII ECLMLSTGELPVPNSRGHTFVHDPCVAGETETGQILRKRTKLAAGDVCIRSFQLT QKAKMFEKALIESVLQITINPTGKVLSTRCAMDEIRIPLMVGSAAVLENTVFHQ DESNMPLDDSTLKKEFDIFSARVYKALEVILKHKDQOEKTRFLKLENTQVW DOAKLHEINADQDEKSDAKSOMEOLKEICGERELTQWETSILDELRLQOIIDIK ATESTYLTQOHEKTLAISENEDTDELMEMORFPERIALETKISKVRMDDEA SYSVLSQNSLETHEYGKLOAEADAHITMKHEDSOLIKCTHNGAPVEHEFTND VAMNLTNRKARLSLENDLDRKKSNEDDLVLMKHLKLNARYSEYDQIOSKIES MSGILRRKDEKERDAEVELSEKNSLRIDERRHQOIVERTLAAGEDVSIIS QKREYLSQKIKVILREKDIINRNADERKGLKLDALSSDKNEIYNERKDKIK KVRGNPNPEKDKKLEINQAFVYDEVENILRSKSOAEOLFSTQSVDAEEOULT KLKRDMAKRFETDSKLOSTIOITSAANDMPKVLQDAMNKEDEKRLFNENGRREMI AFPHLKNHNVCCCEBAFTPDEDEYKQKQKQNSTARSKALMESNAALRQAV QLDLPTITDYAVKLVLEETVPLAKNMLSHLADSSQKQAFDDLLGLAHVQMDRAV EALLQPTITDRHVHEIDQVLEVEDLEYALDSSGRVKSLEELVLENTLQRTDPT IVERDQDLODRHMLNEDMSAQQVWNNHAREEVYASLSLEPQKSEBELVILAEKEE LIVKRLLEESLDPLSEKESLSLEOVYNAKLQDEYHQLAEKREFQEOEDLAGRLKE MKITGYLDSKNEKLEKELQGRHVLCISQSCAKKQOISAEINKSKLQLOGOLK NIDNLTNRKTAQVLEQDTRDISELSEILTSIGLSALEADLKHOSKEKLESEPN WQGLSVYQNSISKQKELKSYQKDEIKRTNPLQULTEMAKMDLDRTYTLDRK LMRHSKMEIEINKIKEIMQOYRTGODIISINSDEGAGTSRYSVVMQGDAA LEMGRSAGQKVLASILIRLAETFCNLGIDIALDEPTNLDPNENESIAAALLRLR MEAKGQENPLIYITHTDERFAHLIGQQLAEKYRYVSKDENQHSIIESQELP"				
BASE COUNT	1537 a	831 c	1061 g	1063 t	
ORIGIN					
Query Match	100.0%; Score 4492; DB 9; Length 4492;				
Best Local Similarity	100.0%; Pred. No. 0;				
Matches 4492; Conservative	0; Mismatches	0; Indels	0; Gaps	0;	
Y	1 aattcgacagagtgatcatcattagcaccatagcgcgtacacaaacctaaagaccctaac	60			
Db	1 AATTCGGCAGCGTGGATCATTTAGACCCATTAGCCGATAGCAAAACCTTAAGAACCTTAC	60			
Y	61 cgggtacaaaccttaaaaaacctaaacgcgcgtggaagactccaaaaaaggagatttttc	120			
Db	61 CGGTACAAACCTTAAMAAACCTTAACGCCGCGGAGAGATCCCAAAAAAGCAGATTTTTC	120			
Y	121 ctcacatgcctcctcctttctcttccaaatcgttttgcaatcactacagcgctaatgaat	180			
Db	121 CTCACATGCCCTCCTCTTCTCTTTCCATCTGTTTGGCAATCATCTACGAGCGTAAGAT	180			
Y	181 agaaagtgtatagagatagcatccgcaactctaggtttgaggcaatcgctctggccagac	240			
Db	181 AGAAGTTGATAGGAGAAATAGCATCCGCAATCTAAGTTTGGGCAATCCGCTGGCGACAG	240			
Y	241 tggatcggaagtgcaagctctaaaggagacattggggctcgttggggcaagatggacac	300			

Db	241	TGgATCGGAATGCAACTGCTGTAAGGAGGACCACTTGGGGCTGTGGGGCAAGATGAGACC	300
QY	301	gttgcacgaagtgtctgtatccaagggatctcgagcttcgatccgacataaagaacytcatc	360
Db	301	GTTCACAAAGATGCTGATCAAGGGGATTCGAGACTTCGATCCGGCACAAATTAACAGTGTATC	360
QY	361	acctcttcaagccgtctcaacctcactcgtttgcccccaagtgctcgtggaagaaccagatc	420
Db	361	ACCTTCCTCAAGCCGCTCACCCCTCATCTGTGGCCCAACGGGtGCTGGGAAACCAACATC	420
QY	421	atcgagtgctgaagcttctcttgcaaccggtcgagctgcgccccaatctccgctctgagcac	480
Db	421	ATCGAGTGCCTGAAGCTTCTTCTGCACCGGCGAGtGtGCCCCCAATCCCGCTtGGGCAC	480
QY	481	acctctgtccacgaccccccaagtagctgtgcgagacgagaaacaaaagacaataaagtgtg	540
Db	481	ACCTTCGTCCACGACCCCAAGGTAGCTGGCGAGACGAGAAACAAAAGGCAAAATTAACTTG	540
QY	541	cggtttaagactgcagcagagaagaagtggtgtgtgcatccgttccttcagcttaaccga	600
Db	541	CGGTTTAAGACTGCAGAGAGAAAGATGtGtGTGTCATCCGCTCTTCACCTTAACCCA	600
QY	601	aaggtcacaagaatgagagtttaaggaatctgaagaagcttcctcaacatataatccacac	660
Db	601	AAGGCTATCAAAAGATGAGGTTTAAGGCCAATTGAAAGGtGTCCTCCACACTATAATCCACAC	660
QY	661	acaaagggaagaagtctgcctcaagctacagatgtctgacatgatatagagagatccgtcc	720
Db	661	ACAGGGGAGAAAGTCTGCTCCTCACACTCAAGATGtGTCATGATGATGAGAGATTCCTGGC	720
QY	721	ttaatgggtgttctgaaagccgtactcggaggaatgttatcttgctcaccagaatagaatcc	780
Db	721	TTAATGGGTGTTCGAAAGGCCGTACTGGAGAAAGTAAATTGTTCACCAAGATGAATCC	780
QY	781	aattggccatgtgcagagcccgctcaaacacttaagaagaagttcgaagacatctctctgcc	840
Db	781	AATTGGCATTGCGAGACCCGCTCACACTTAAGAAATGTAAGAAATGTAAGATCTGTCTGGCC	840
QY	841	acacgctatacgaagaagctcttgaagatcaataaagaacactcacagagatcoaatgtcaag	900
Db	841	ACACGCTATACGAAGAAGCTCTTGAAGTCAATTAAGAAACTTCACACAGATCAATGTCAAG	900
QY	901	atcaagacttttagttaaagctggggaaccttcagaactgtataaagacaaagacataaag	960
Db	901	ATCAAGACTTTTAACTTAAAGCTGGAGAACCTTCAACACTGTAAAGACCAACACATTAAG	960
QY	961	ctgcgttgaaaatctgtctcaagaatcaagaagaagctcagatcgtcccaaatctcaagttag	1020
Db	961	CTGCCTGAAAATATTGCTCAAGATCAAGAAATCATCAAGTCCCTCAAAATCTCAGATGAG	1020
QY	1021	caactgaagagaagaatctgtgtgtaccgagagagaatccctgcgaatctgnaacaagtgtg	1080
Db	1021	CAACTGAAGGAAAGATCTGTGTAACGAGAGGAATCTCTCAAAATGGAACAAAGTTTG	1080
QY	1081	gatgaactcgagaagaacttcgaggacaatattgacataaagacaagagagaaggtacetta	1140
Db	1081	GATGAACGAAAMAACTTCAGGGCAAAATTGAATTAAGGCAACAGAGGAAGTACATTA	1140
QY	1141	cttaacgagcagcaatgaagaagctgtgcaccttctcgagaagaatctgaagatccgattag	1200
Db	1141	CTTAGCGAGCAAGCATGAAAAGCTTCGTCGCACTTTCGAGGAAGAAAGATATCCGATGAG	1200
QY	1201	gaactaatggaatgycgaacaaaatttgaaagaaggatgcygttactagaaaacaaaatc	1260
Db	1201	GAACTAAAGGATGGCAAAACAAATTTGAGAGAAAGATTTGCGTTATCTAATAACAAAAATC	1260
QY	1261	agtaactgttaagaagatgatgatgatgaagacacttatagctccgcttctgtccaaaca	1320
Db	1261	AGTAACCTTGTAAAGATGTGATGATGAAGCAATCTTAATAGCTCCGTTCTGTCCAAACAA	1320
QY	1321	aactctgaattacaacatgaaatctgaaagctccagagcagaagctgattgtctcaactgact	1380
Db	1321	AATTCGAAATTAACATGAATTTGGAAGCTCCAGGACGAAGGCTGATGCTCACTGACT	1380

QY 1381 atgaaacatgaaacgactcagacataaaaaaataatgactaaacataacttggccg 1440
DB 1381 atgaaacatgaaacgactcagacataaaaaaataatgactaaacataacttggccg 1440
QY 1441 gtctctgaaacaccccttgaagatgctgtaagaacctcaaaacagatgaaagc 1500
DB 1441 gtctctgaaacaccccttgaagatgctgtaagaacctcaaaacagatgaaagc 1500
QY 1501 agacatcaagctctggaagatgctgtaagaagaaatccaataagaatgctta 1560
DB 1501 agacatcaagctctggaagatgctgtaagaagaaatccaataagaatgctta 1560
QY 1561 gatgtcttggaacacactatcttaaaataaagtctgactccgaagctgtagcag 1620
DB 1561 gatgtcttggaacacactatcttaaaataaagtctgactccgaagctgtagcag 1620
QY 1621 atacaactaagatgaaatccaatgcaagcatttaagaaggagaataaagaaga 1680
DB 1621 atacaactaagatgaaatccaatgcaagcatttaagaaggagaataaagaaga 1680
QY 1681 gaacggatgctggaagtgagcttccaataatccatccgtaatgagag 1740
DB 1681 gaacggatgctggaagtgagcttccaataatccatccgtaatgagag 1740
QY 1741 gagagacatgcaaatggaagtcgagaggaagacactgagcttggaagagagact 1800
DB 1741 gagagacatgcaaatggaagtcgagaggaagacactgagcttggaagagagact 1800
QY 1801 gattcaattatagtcagaaacgaaacagatataatgcttggaacagaaataaagtg 1860
DB 1801 gattcaattatagtcagaaacgaaacagatataatgcttggaacagaaataaagtg 1860
QY 1861 ctcttgcggaagaaagataataataatgaaatgctgtaagaagtaaaactggcttg 1920
DB 1861 ctcttgcggaagaaagataataataatgaaatgctgtaagaagtaaaactggcttg 1920
QY 1921 aagaagatgcatggaagcagcaagcagcaagcctcaatgagatgtaatgagataag 1980
DB 1921 aagaagatgcatggaagcagcaagcagcaagcctcaatgagatgtaatgagataag 1980
QY 1981 gataaaatcaaaaagtaacttaagggaagaaaccccttggaagaaatataagaagag 2040
DB 1981 gataaaatcaaaaagtaacttaagggaagaaaccccttggaagaaatataagaagag 2040
QY 2041 atcaatcaagcccttgccgtgagcaaggaatacaatgagtaagaatccaatccag 2100
DB 2041 atcaatcaagcccttgccgtgagcaaggaatacaatgagtaagaatccaatccag 2100
QY 2101 gaagcagagcagaagcttaaaatttactcagagcagaagtaactgctcagaagaacttg 2160
DB 2101 gaagcagagcagaagcttaaaatttactcagagcagaagtaactgctcagaagaacttg 2160
QY 2161 acaaaactcgaagagatalatgaaatcaaaaagaagatccctggaactcgaactcaatc 2220
DB 2161 acaaaactcgaagagatalatgaaatcaaaaagaagatccctggaactcgaactcaatc 2220
QY 2221 atttacaagatctgcttaagtgtgaacatgcttcccaagatctcaaaagcgcgaagac 2280
DB 2221 atttacaagatctgcttaagtgtgaacatgcttcccaagatctcaaaagcgcgaagac 2280
QY 2281 aaagaagatcgaagaaagatgagaatctgcaaatggaatgagcgggaatgcttga 2340
DB 2281 aaagaagatcgaagaaagatgagaatctgcaaatggaatgagcgggaatgcttga 2340
QY 2341 cctttgaaacatctgctggaagaaatcatgtaacatgctgtaagctgcttca 2400
DB 2341 cctttgaaacatctgctggaagaaatcatgtaacatgctgtaagctgcttca 2400
QY 2401 cctgagatgagagatgagctgctgtaagaagaagaagatgcaaaactcaagtaagcag 2460
DB 2401 cctgagatgagagatgagctgctgtaagaagaagaagatgcaaaactcaagtaagcag 2460

QY 2461 agatcctaaagctctgcaaatggaatcatcaaatgctggaagccttttcaggcaattgat 2520
DB 2461 agatcctaaagctctgcaaatggaatcatcaaatgctggaagccttttcaggcaattgat 2520
QY 2521 aaactcgaactactatgactgcttaagtgtgtaagaagaaacacactacactagca 2580
DB 2521 aaactcgaactactatgactgcttaagtgtgtaagaagaaacacactacactagca 2580
QY 2581 gagaagaaacttgaactaaatcttgccggaatggaagtcaagaagcgaagcatttgat 2640
DB 2581 gagaagaaacttgaactaaatcttgccggaatggaagtcaagaagcgaagcatttgat 2640
QY 2641 ctttggtgctctgcccattcgaatggaagcaggaatgagcagtggaagccttataca 2700
DB 2641 ctttggtgctctgcccattcgaatggaagcaggaatgagcagtggaagccttataca 2700
QY 2701 ccactgataactatgacagcagcagtlacatgaataacacagactagcacaagaatgaa 2760
DB 2701 ccactgataactatgacagcagcagtlacatgaataacacagactagcacaagaatgaa 2760
QY 2761 gatctggaatgacactgactgtagtgccgaagtgctcaagcttggaggaataatca 2820
DB 2761 gatctggaatgacactgactgtagtgccgaagtgctcaagcttggaggaataatca 2820
QY 2821 ctggaagctgaactctctgcaagaaacagagcacaatgctgctgaagtgatgactct 2880
DB 2821 ctggaagctgaactctctgcaagaaacagagcacaatgctgctgaagtgatgactct 2880
QY 2881 agagatcaacatagaatgcttaaatgaagatgctcaagtgctcagatgagatgacat 2940
DB 2881 agagatcaacatagaatgcttaaatgaagatgctcaagtgctcagatgagatgacat 2940
QY 2941 gctcgggaagagaagtgaaagctctcagcatatggaataatccaaaactcgaagag 3000
DB 2941 gctcgggaagagaagtgaaagctctcagcatatggaataatccaaaactcgaagag 3000
QY 3001 gaattggtgctctgtagcaggaagaaacacacactgctgagaagaagacctttaga 3060
DB 3001 gaattggtgctctgtagcaggaagaaacacacactgctgagaagaagacctttaga 3060
QY 3061 gactccttgatccatctgccaagaagaaagagcctgttgcaagagatataatgcttg 3120
DB 3061 gactccttgatccatctgccaagaagaaagagcctgttgcaagagatataatgcttg 3120
QY 3121 aagcaaaagctggaatgaaagatatacaagcttgcaagaagaagaagagcttcagca 3180
DB 3121 aagcaaaagctggaatgaaagatatacaagcttgcaagaagaagaagagcttcagca 3180
QY 3181 gaacttgatgctctggaagaaacttaataatgaagataaagaagctctggaatccaaga 3240
DB 3181 gaacttgatgctctggaagaaacttaataatgaagataaagaagctctggaatccaaga 3240
QY 3241 aacgaaagcttgaagaaatgcaaggaagcagcttcttcgaatccagatcagag 3300
DB 3241 aacgaaagcttgaagaaatgcaaggaagcagcttcttcgaatccagatcagag 3300
QY 3301 tgcattgcaaaaacagcaaaagatatacagctgagtttaaaacagacaaagaaactgacag 3360
DB 3301 tgcattgcaaaaacagcaaaagatatacagctgagtttaaaacagacaaagaaactgacag 3360
QY 3361 ggcgaagcgcagcttggaagaagaacatggaatcaaatcgaagtcaggaagaagaagct 3420
DB 3361 ggcgaagcgcagcttggaagaagaacatggaatcaaatcgaagtcaggaagaagaagct 3420
QY 3421 gatgtgaaacaaactactcgtgataagaatcaccttgaaagaagcgtgcttcaatagct 3480
DB 3421 gatgtgaaacaaactactcgtgataagaatcaccttgaaagaagcgtgcttcaatagct 3480
QY 3481 agcttgctgctatagaagctgactcgaagcgcacatctcaagaagaagaagagctta 3540
DB 3481 agcttgctgctatagaagctgactcgaagcgcacatctcaagaagaagaagagctta 3540
QY 3541 tcagaatttaacagtggaagaaacactctcttatacaagaatataattccaagcac 3600

```
|||||
Db 3541 TCAGATTTTACAGGCGCAGAGCACTTCTGTTATCAAGTAATTTCAAGCAC 3600
|||
Qy 3601 aacaaagagcttaactgacagatcacagagatcacagagcgatatactaatcatt 3660
|||||
Db 3601 AAACAAGAGCTTAACTGACAGATCAAGATATCGAAGACGATATCTAATCAATTT 3660
|||||
Qy 3661 cccaaacttaagacaactgaatgaatgcaaaacagagacttgacagatatatactgctta 3720
|||||
Db 3661 CTCACACTTAAGCAACTGAATGGCAAAACAGACTTGACAGATATATACTGCTTA 3720
|||||
Qy 3721 gacaaggctctatgcygttccacagcatgaagatgagagagataataataatcaag 3780
|||||
Db 3721 GACAAGGCTCTTATCGGTTCCACACATGAAGATGAGAGATTAATAATAATCAAG 3780
|||||
Qy 3781 gaactgtgcaacagacacacagagagcgagatattgattacataagcaataattctgat 3840
|||||
Db 3781 GAACGTGGCAACAGACATACAGAGCGCAGGATATGATTTACATTAAGCATTAATTTCTGAT 3840
|||||
Qy 3841 tctgaggtgctgacacgcacatcacacacacacacacacacacacacacacacacacac 3900
|||||
Db 3841 TCTGAGGCTGCTGCGCATGCTGATCATACGCGGTGTGTTATGCAAACTGGTATGCT 3900
|||||
Qy 3901 gacgtggaatgcgagggcgctgcagtgctgtcagaaggttctctctctataatc 3960
|||||
Db 3901 GACGTGGAATGCGAGGCGCTGCAGTGTGTCAGAAAGTTCTTCTCTTATATATC 3960
|||||
Qy 3961 agactgacacttgaggaaacttctcctgaactgcgtatattgcttggatgagcca 4020
|||||
Db 3961 AGACTGACACTTGGGAAACTTCTGCTGAATCGCGTATATGGCTTTGGATGAGCCA 4020
|||||
Qy 4021 aactgaatcagatgagcgaatgcaagagatctgctgctgctgctgctgctgctgctgctg 4080
|||||
Db 4021 ACTAGCATATAGATGGCCAAATGCGAGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4080
|||||
Qy 4081 gaagcagagaaagggcagaggaactccagttcattgttaactcaactatgtagagattc 4140
|||||
Db 4081 GAAGCCAGGAAAGGCGCAGGAACTTCCAGTTATTTATCACTCATGATGAGAGATTT 4140
|||||
Qy 4141 gcccatcttctgctcaagagcagctgctgagagatcagatcagatcagatcagatcagatcag 4200
|||||
Db 4141 GCCCATCTTATCGGTCAAGGCAAGCTTGTGAGATGATCTATGAGATCTCCAAAGATGAG 4200
|||||
Qy 4201 aacagacagacataatgaatcccaagagataattgaactaagggtgtctcagagagctg 4260
|||||
Db 4201 AACAGACAGACATTAATGATCCCAAGAGATATTGACTAAGGCTGTCTAGAGAGGCTG 4260
|||||
Qy 4261 tagcagcactcgttctgctagtcagatccagttcaattatgccaagatcgtgtgcagag 4320
|||||
Db 4261 TAGCAGCCTCGTTGCTAGTCCATCCAGTTAATTTATGCCAAGTACTGTCAGAG 4320
|||||
Qy 4321 caatgttaacaagctttaagaggtcgtttagcgcttaagctgctgctgctgctgctgctgctg 4380
|||||
Db 4321 CAAAGTTAAACAAGCTTAAAGAGGCTGCTGTTACGCGCTTACGTCAGTTCGTTGANAACCTAT 4380
|||||
Qy 4381 ccttcgttctgtatacttaattatctgcccagagagatgaatgtgtgactgtgtgactgtgtg 4440
|||||
Db 4381 CCTTCGTTGTTGTTATTTATTTATCTGGCCAGAGATGGAATGTGTGCACTGGTGAT 4440
|||||
Qy 4441 ggaatgttcagacataatgaatgtttcacaatctagcatcaaaaaa 4492
|||||
Db 4441 GGATGTTCAAGCATCAATGAATGATTTCACAATCTAGCATCAAAAAA 4492
|||||
```

Search completed: November 8, 2001, 00:52:23
Job time: 42833 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd

OM nucleic - nucleic search, using sw model

Run on: November 7, 2001, 22:20:31 ; Search time:260.71 seconds
(without alignments)
10818.665 Million cell updates/sec

Title: US-09-538-396-1

Sequence: 1 aattcgacagtgatcc.....atctagcatcaaaaaaaaa 4492

Scoring table: OLIgo_NUC Gapon 60.0 Gapext 60.0

Searched: 730101 seqs, 313950809 residues

word size : 30

Total number of hits satisfying chosen parameters: 1

Minimum	DB	seq	length:	0
---------	----	-----	---------	---

Maximum DB seq length: 200000000000

Post-processing: Listing first 45 summaries

```
Database :
_N_Geneseq_0601:*
1:/SIDSL/gcgdata/geneseq/geneseqn/NA1980.DAT.*
2:/SIDSL/gcgdata/geneseq/geneseqn/NA1981.DAT.*
3:/SIDSL/gcgdata/geneseq/geneseqn/NA1982.DAT.*
4:/SIDSL/gcgdata/geneseq/geneseqn/NA1983.DAT.*
5:/SIDSL/gcgdata/geneseq/geneseqn/NA1984.DAT.*
6:/SIDSL/gcgdata/geneseq/geneseqn/NA1985.DAT.*
7:/SIDSL/gcgdata/geneseq/geneseqn/NA1986.DAT.*
8:/SIDSL/gcgdata/geneseq/geneseqn/NA1987.DAT.*
9:/SIDSL/gcgdata/geneseq/geneseqn/NA1988.DAT.*
10:/SIDSL/gcgdata/geneseq/geneseqn/NA1989.DAT.*
11:/SIDSL/gcgdata/geneseq/geneseqn/NA1990.DAT.*
12:/SIDSL/gcgdata/geneseq/geneseqn/NA1991.DAT.*
13:/SIDSL/gcgdata/geneseq/geneseqn/NA1992.DAT.*
14:/SIDSL/gcgdata/geneseq/geneseqn/NA1993.DAT.*
15:/SIDSL/gcgdata/geneseq/geneseqn/NA1994.DAT.*
16:/SIDSL/gcgdata/geneseq/geneseqn/NA1995.DAT.*
17:/SIDSL/gcgdata/geneseq/geneseqn/NA1996.DAT.*
18:/SIDSL/gcgdata/geneseq/geneseqn/NA1997.DAT.*
19:/SIDSL/gcgdata/geneseq/geneseqn/NA1998.DAT.*
20:/SIDSL/gcgdata/geneseq/geneseqn/NA1999.DAT.*
21:/SIDSL/gcgdata/geneseq/geneseqn/NA2000.DAT.*
22:/SIDSL/gcgdata/geneseq/geneseqn/NA2001.DAT.*
```

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	4492	100.0	4492	22	AAC66915	Maize RMD50 coding

```

RESULT      1
AAC66915
ID      AAC66915 standard; cDNA; 4492 BP
XX
AC      AAC66915;

```

XX		27-MAR-2001	(first entry)	
DT				
XX				
DE		Maize RAD50 coding sequence.		
XX				
KW		Maize; RAD50; meiotic recombination; DNA repair; chromosome 4p;		
XX		recombination efficiency; transformation efficiency; ss.		
OS		Zea mays.		
XX				
PN		WO200068404-A1.		
XX				
PD		16-NOV-2000.		
XX				
PF		25-APR-2000; 2000MO-US11086.		
XX				
PR		05-MAY-1999; 99US-0132575.		
XX				
PA		(PION-) PIONEER HI-BRED INT INC.		
PI		Mahajan PB, Shi J;		
XX				
DR		WPI; 2001-007402/01.		
DR		P-PSTDB; AAB27248.		
PT		Nucleic acids encoding Rad50 polypeptides, useful for transgenic plant		
PT		production and for producing antibodies that are useful for modifying		
XX		gene expression -		
PS				
PS		Claim 1; Page 65-70; 76pp; English.		
XX				
CC		The present sequence is the coding sequence of the maize RAD50 gene,		
CC		found on maize chromosome 4p. RAD50 is involved in meiotic recombination		
CC		and DNA repair, and the gene and its protein can be used to produce		
CC		transgenic plants whose expression of the gene may be regulated. This can		
CC		be useful in the regulation of transformation and recombination		
CC		efficiency in plants.		
SO				
XX		Sequence 4492 BP; 1537 A; 831 C; 1061 G; 1063 T; 0 other:		
Query Match	100.0%; Score 4492; DB 22; Length 4492;			
Best Local Similarity	100.0%; Pred. No. 0;			
Matches 4492; Conservative	0; Mismatches 0; Indels 0; Gaps 0;			
OY	1 aattcgcacgagatgtgatccattagcaccccaataaggctacaaaaaccttaagaacctaac	60		
Db	1 aattcgacgaagtgtgattccattagcaccccaataaggctacaaaaaccttaagaacctaac	60		
OY	61 cgggtacaaaaccctaataaaaccctaagcgcgtgaagactccaanaaaacgcgatttttc	120		
Db	61 cgggtacaaaaccctaataaaaccctaagcgcgtgaagactccaanaaaacgcgatttttc	120		
OY	121 ctccactgccctccttcttctctttccacaatcgltttgcaatcactacgaagcgtaatgaat	180		
Db	121 ctccactgccctccttcttctctttccacaatcgltttgcaatcactacgaagcgtaatgaat	180		
OY	161 agaaatttataggagatatagcatccgcgaatctcaagttttggggcaatcgcttcggccaagc	240		
Db	161 agaaatttataggagatatagcatccgcgaatctcaagttttggggcaatcgcttcggccaagc	240		
OY	241 tggatcgcgaagtccaagtcgttagaagggaagcatttgsggtctgttggggcaagatgagcac	300		
Db	241 tggatcgcgaagtccaagtcgttagaagggaagcatttgsggtctgttggggcaagatgagcac	300		
OY	301 gttgacaagatgctgataccaagggaatccggaagcttcogataccggaacaataagaagctcatc	360		
Db	301 gttgacaagatgctgataccaagggaatccggaagcttcogataccggaacaataagaagctcatc	360		
OY	361 acctcttcacaagccgctcacctcctatcgcttgcgcccaacgaagcgctggcaagaccagatc	420		
Db	361 acctcttcacaagccgctcacctcctatcgcttgcgcccaacgaagcgctggcaagaccagatc	420		

421 atcgatgctcgaagcttcttctgcaacgagctgcccccaactccgcctctgacac 480
421 atcgatgctcgaagcttcttctgcaacgagctgcccccaactccgcctctgacac 480
481 accctcgcacagaccccaagctgctgcgagacgagacaacaagaagacaattagtg 540
481 accctcgcacagaccccaagctgctgcgagacgagacaacaagaagacaattagtg 540
541 cggtttaagactcgcagagagaaagatgctgctgcatccgctctccagcttccaa 600
541 cggtttaagactcgcagagagaaagatgctgctgcatccgctctccagcttccaa 600
601 aagacatcaaaatgagatgtaaggcaattgaaagctccctccaaactataatccac 660
601 aagacatcaaaatgagatgtaaggcaattgaaagctccctccaaactataatccac 660
661 aagagagagaaagctcgcctcagctcaacagatgctgacatgataagagatccctgc 720
661 aagagagagaaagctcgcctcagctcaacagatgctgacatgataagagatccctgc 720
721 ttaatgggtgttcgaaagcgctactgagaaatgtatattgttcccaagaatgaacc 780
721 ttaatgggtgttcgaaagcgctactgagaaatgtatattgttcccaagaatgaacc 780
781 aattggcatcgcagagacccgctcaacacttaagaagaagttcgatgacatctctgc 840
781 aattggcatcgcagagacccgctcaacacttaagaagaagttcgatgacatctctgc 840
841 aagagatcaaaagctcttgaaagcttaagaagaagaactcccaagaatcaatgcaag 900
841 aagagatcaaaagctcttgaaagcttaagaagaagaactcccaagaatcaatgcaag 900
901 atcaaaacttttaggttaaaagcttgagaaacttcagactgttaaaagacagacaag 960
901 atcaaaacttttaggttaaaagcttgagaaacttcagactgttaaaagacagacaag 960
961 ctgcgtgaaataatctctcaagatcaagaagaagtcagctccaaactccagatggag 1020
961 ctgcgtgaaataatctctcaagatcaagaagaagtcagctccaaactccagatggag 1020
1021 caactgaagagaaagctcgtgctgacagagagaaactccgcaatgggaaacattg 1080
1021 caactgaagagaaagctcgtgctgacagagagaaactccgcaatgggaaacattg 1080
1081 gatgaactcgaagaagcttccaggaacttgacatcaagcaacagagagaatatact 1140
1081 gatgaactcgaagaagcttccaggaacttgacatcaagcaacagagagaatatact 1140
1141 cttaagcagcagacatgaaagctgctgcacattctcgaagaaatgaaagctacagatg 1200
1141 cttaagcagcagacatgaaagctgctgcacattctcgaagaaatgaaagctacagatg 1200
1201 gaactaagagagagagaaacttgaaagagagatgctgactacagaaacaaatc 1260
1201 gaactaagagagagagaaacttgaaagagagatgctgactacagaaacaaatc 1260
1261 agcaaaacttgaaagagatgataagctgataagcatctatagctcgttctccaaacaa 1320
1261 agcaaaacttgaaagagatgataagctgataagcatctatagctcgttctccaaacaa 1320
1321 aattcgaatcaacaatgaaatgaaagctccagagcaaaagctgctcagctgact 1380
1321 aattcgaatcaacaatgaaatgaaagctccagagcaaaagctgctcagctgact 1380
1381 atgaagaatgaagaagactcagacataaaatataatgacataaactaactctggcgcg 1440
1381 atgaagaatgaagaagactcagacataaaatataatgacataaactaactctggcgcg 1440
1441 gtctcgaacatcccttcaagaaatgctgctgataagacatcaaaacagatgaaagcg 1500
1441 gtctcgaacatcccttcaagaaatgctgctgataagacatcaaaacagatgaaagcg 1500
1501 agactacaaagctctgagaatgctgctgataagaagaatcccaatgaagatcaatga 1560
1501 agactacaaagctctgagaatgctgctgataagaagaatcccaatgaagatcaatga 1560

1501 agactacaaagctctgagaatgctgctgataagaagaatcccaatgaagatcaatga 1560
1501 agactacaaagctctgagaatgctgctgataagaagaatcccaatgaagatcaatga 1560
1561 gatgtttgtgaaacactatctttaaataatgctgctgactccgaagtgtatgctcag 1620
1561 gatgtttgtgaaacactatctttaaataatgctgctgactccgaagtgtatgctcag 1620
1621 atacaatcaagatgaaatccatgctcagacatcttaagaacgagaaagatataagaaga 1680
1621 atacaatcaagatgaaatccatgctcagacatcttaagaacgagaaagatataagaaga 1680
1681 gaaacgagatgctgcagaaagtggagcttcaaaaattaatctatcccgtaatgaaag 1740
1681 gaaacgagatgctgcagaaagtggagcttcaaaaattaatctatcccgtaatgaaag 1740
1741 gaaagacatgcaaatgaaatgaaagcagagacatctgctgagaaagaaactat 1800
1741 gaaagacatgcaaatgaaatgaaagcagagacatctgctgagaaagaaactat 1800
1801 gatcaatataagtcagaaacgagacagatataatgcttgaaacagaaataaagtg 1860
1801 gatcaatataagtcagaaacgagacagatataatgcttgaaacagaaataaagtg 1860
1861 ctctcgcgagaaagataataataagaaatgctgataagaaagaaactgggttg 1920
1861 ctctcgcgagaaagataataataagaaatgctgataagaaagaaactgggttg 1920
1921 aagaagatgcatctgaaagcagacagacagacatcaatgaaatgataatgactaag 1980
1921 aagaagatgcatctgaaagcagacagacagacatcaatgaaatgataatgactaag 1980
1981 gataaataaagaagctttagtgagagaaactcttgaaagagataatgaagaagag 2040
1981 gataaataaagaagctttagtgagagaaactcttgaaagagataatgaagaagag 2040
2041 atcaatcaagccttctgagctctgcaagaaatcaaatgaaatgaaatccag 2100
2041 atcaatcaagccttctgagctctgcaagaaatcaaatgaaatgaaatccag 2100
2101 gaagcagagcaagaagcttaattactcagagcaagaagtaactgctgagaaacattg 2160
2101 gaagcagagcaagaagcttaattactcagagcaagaagtaactgctgagaaacattg 2160
2161 acaaaactcgaagagatgctgcaagaaagaaatccctgagacccgaactcaatc 2220
2161 acaaaactcgaagagatgctgcaagaaagaaatccctgagacccgaactcaatc 2220
2221 atttacaatatctgctaaatgctgcaatgcttccaaagttctacaagaacgacatgac 2280
2221 atttacaatatctgctaaatgctgcaatgcttccaaagttctacaagaacgacatgac 2280
2281 aaaaagatgaagaaagaagatagaagatctccgaagaaatgctgagaaatgctgca 2340
2281 aaaaagatgaagaaagaagatagaagatctccgaagaaatgctgagaaatgctgca 2340
2341 cctttgagacatctgctcgaagaaatcagatgctcccaagctgtaagctgcttccaa 2400
2341 cctttgagacatctgctcgaagaaatcagatgctcccaagctgtaagctgcttccaa 2400
2401 cctgagagagagatgagctgctgagaaagaaagatgcaaaactcaatgactgagag 2460
2401 cctgagagagagatgagctgctgagaaagaaagatgcaaaactcaatgactgagag 2460
2461 agactcaaaagctcctgcaagaaatcagaaatgctgaaagctcttctccgaactggat 2520
2461 agactcaaaagctcctgcaagaaatcagaaatgctgaaagctcttctccgaactggat 2520
2521 aaactcggacatctcagatgctgtaagctgtaagaaagaaacataactcagca 2580
2521 aaactcggacatctcagatgctgtaagctgtaagaaagaaacataactcagca 2580
2581 gaaagaaactgaaatcaaacattgctgagaaagaaagctgcaagcattgagat 2640
2581 gaaagaaactgaaatcaaacattgctgagaaagaaagctgcaagcattgagat 2640

Dh 2581 gagaaaaacttgatcaaacatttggcggatgaagtcgaagcgcgagcatttgat 2640
Qy 2641 cttttgggtgtcttgcacatgttcaaatggaaggaagcaggttgaaacctattaca 2700
Dh 2641 cttttgggtgtcttgcacatgttcaaatggaaggaagcaggttgaaacctattaca 2700
Qy 2701 cccactgtacctatgacagcgcatgtacatgaatccaacaaactgaatgaagtagaa 2760
Dh 2701 cccactgtacctatgacagcgcatgtacatgaatccaacaaactgaatgaagtagaa 2760
Qy 2761 gattctgaataltgacatttgatctctgtgcccgaagtgatcaagcttttgaggaatcaa 2820
Dh 2761 gattctgaataltgacatttgatctctgtgcccgaagtgatcaagcttttgaggaatcaa 2820
Qy 2821 ctggagcggaactcttctcgaaggaacaaaggaacacattgttgcgaagtgaatcctt 2880
Dh 2821 ctggagcggaactcttctcgaaggaacaaaggaacacattgttgcgaagtgaatcctt 2880
Qy 2881 agagatcaacatagatgtctaaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 2940
Dh 2881 agagatcaacatagatgtctaaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 2940
Qy 2941 gctcggagagagaaagtgaaagcttctagcatattggaagaatcccaaaatctgaaga 3000
Dh 2941 gctcggagagagaaagtgaaagcttctagcatattggaagaatcccaaaatctgaaga 3000
Qy 3001 gaattgggcttcttctcgaaggaacaaacacatgattgtcgaagaagaagcttttagaa 3060
Dh 3001 gaattgggcttcttctcgaaggaacaaacacatgattgtcgaagaagaagcttttagaa 3060
Qy 3061 gaattgggcttcttctcgaaggaacaaacacatgattgtcgaagaagaagcttttagaa 3120
Dh 3061 gaattgggcttcttctcgaaggaacaaacacatgattgtcgaagaagaagcttttagaa 3120
Qy 3121 aagcaaaagctgagatgaagatcatcatcagcttgcagaagaagaagagatccagcaa 3180
Dh 3121 aagcaaaagctgagatgaagatcatcatcagcttgcagaagaagaagagatccagcaa 3180
Qy 3181 gaactgtgtcttctcgaaggaacaaacacatgattgtcgaagaagaagcttttagaa 3240
Dh 3181 gaactgtgtcttctcgaaggaacaaacacatgattgtcgaagaagaagcttttagaa 3240
Qy 3241 aacgaaagcttaaggaattgcaaggaagcgttcttgcattccaaagttagaagct 3300
Dh 3241 aacgaaagcttaaggaattgcaaggaagcgttcttgcattccaaagttagaagct 3300
Qy 3301 tgcattgcaaaaacagcaaaagatcatcagcttgaatgaagaagaagaagacttgcag 3360
Dh 3301 tgcattgcaaaaacagcaaaagatcatcagcttgaatgaagaagaagaagacttgcag 3360
Qy 3361 ggcgaagcgcaagtgcaaaaacatgtatgacaatctcaagtagaagaaacaaagct 3420
Dh 3361 ggcgaagcgcaagtgcaaaaacatgtatgacaatctcaagtagaagaaacaaagct 3420
Qy 3421 gatctggaacacatcttctgcatatgaatcaacttgaagaagaagcttcttcaatagct 3480
Dh 3421 gatctggaacacatcttctgcatatgaatcaacttgaagaagaagcttcttcaatagct 3480
Qy 3481 agcttgcgtcatagaaagcgtatcgaacgcattctcaagaagaagaagagcttaac 3540
Dh 3481 agcttgcgtcatagaaagcgtatcgaacgcattctcaagaagaagaagagcttaac 3540
Qy 3541 tcaagaatttaacagatggaagaacacattcttcttcaagaagaagaagac 3600
Dh 3541 tcaagaatttaacagatggaagaacacattcttcttcaagaagaagaagac 3600
Qy 3601 aaacaagaagcttaacatgcatatgaagaatctgaagaagcgaatataactaactt 3660
Dh 3601 aaacaagaagcttaacatgcatatgaagaatctgaagaagcgaatataactaactt 3660
Qy 3661 ctccagcttaagacaactgaatgcaaaaacaaagaaagcttgcagagatatatactt 3720
Dh 3661 ctccagcttaagacaactgaatgcaaaaacaaagaaagcttgcagagatatatactt 3720

Qy 3721 gacaagcctcttatgtcgttccacagcatgaagatggaaggaataaataatcaag 3780
Dh 3721 gacaagcctcttatgtcgttccacagcatgaagatggaaggaataaataatcaag 3780
Qy 3781 gaactgtgcaacacacatacagaagcgaggaattgaatgaatgaatgaatgaatgaat 3840
Dh 3781 gaactgtgcaacacacatacagaagcgaggaattgaatgaatgaatgaatgaatgaat 3840
Qy 3841 tctgaaggtgtctgcaatcgaatcaacacacacacacacacacacacacacacacacac 3900
Dh 3841 tctgaaggtgtctgcaatcgaatcaacacacacacacacacacacacacacacacacac 3900
Qy 3901 gactgtggaatgcaagagcgctgcaagctgctgctgcaagagcttcttcttcaatc 3960
Dh 3901 gactgtggaatgcaagagcgctgcaagctgctgctgcaagagcttcttcttcaatc 3960
Qy 3961 agactgcaactgcaagaaacttctgcacacacacacacacacacacacacacacacacac 4020
Dh 3961 agactgcaactgcaagaaacttctgcacacacacacacacacacacacacacacacacac 4020
Qy 4021 actacgaatctagatggaagcaaatgcaagaagcttctgctgctgcaagagcttctgcaatg 4080
Dh 4021 actacgaatctagatggaagcaaatgcaagaagcttctgctgctgcaagagcttctgcaatg 4080
Qy 4081 gaagcgaagaaagggcaggaagaactccagcttgaatgaatcaactcaatgaagatt 4140
Dh 4081 gaagcgaagaaagggcaggaagaactccagcttgaatgaatcaactcaatgaagatt 4140
Qy 4141 gccatcttatgtcgaaggaagcagcttctggaaggaactcaacagcttccaaagtag 4200
Dh 4141 gccatcttatgtcgaaggaagcagcttctggaaggaactcaacagcttccaaagtag 4200
Qy 4201 aacacgaacagataattgaatcccaagaagatattgaatgaaggggttctcagagagct 4260
Dh 4201 aacacgaacagataattgaatcccaagaagatattgaatgaaggggttctcagagagct 4260
Qy 4261 tagcagcactgttctgctgctgcaatccagcttgaatgaatgaatgaatgaatgaatgaat 4320
Dh 4261 tagcagcactgttctgctgctgcaatccagcttgaatgaatgaatgaatgaatgaatgaat 4320
Qy 4321 caatgttaacaaagcttgaagagcgttctggaaggaactcaacagcttccaaagtag 4380
Dh 4321 caatgttaacaaagcttgaagagcgttctggaaggaactcaacagcttccaaagtag 4380
Qy 4381 ccttgcgttctatcaactatttaactggaaggaaggaaggaaggaaggaaggaaggaaggaag 4440
Dh 4381 ccttgcgttctatcaactatttaactggaaggaaggaaggaaggaaggaaggaaggaaggaag 4440
Qy 4441 ggaatgttcaacagacatgaatgaatgttcaacaaatgaacatcaacaaacaaacaaacaaac 4492
Dh 4441 ggaatgttcaacagacatgaatgaatgttcaacaaatgaacatcaacaaacaaacaaacaaac 4492

Search completed: November 8, 2001, 00:59:38
Job time: 9547 sec

THU NOV 8 09:02:32 2001

us-09-538-396-1.rng

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2001, 12:33:50 ; Search time 3157.09 Seconds
(without alignments)
13449.796 Million cell updates/sec

Title: US-09-538-396-1

Perfect score: 4492
Sequence: 1 aattggcgcagtgatcc.....atctagcatcaaaaaaaaa 4492

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 10228115 seqs, 4726426750 residues

Word size : 30

Total number of hits satisfying chosen parameters: 1

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: gb_est41:*
42: gb_est42:*
43: gb_est43:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_esthum29:*
64: em_esthum30:*
65: em_esthum31:*
66: em_esthum32:*
67: em_esthum33:*
68: em_esthum34:*
69: em_esthum35:*
70: em_esthum36:*
71: em_esthum37:*
72: em_esthum38:*
73: em_esthum39:*
74: em_esthum40:*
75: em_esthum41:*
76: em_esthum42:*
77: em_esthum43:*
78: em_esthum44:*
79: em_esthum45:*
80: em_esthum46:*
81: em_esthum47:*
82: em_esthum48:*
83: em_esthum49:*
84: em_esthum50:*
85: em_esthum51:*
86: em_esthum52:*
87: em_esthum53:*
88: em_esthum54:*
89: em_esthum55:*
90: em_esthum56:*
91: em_esthum57:*
92: em_esthum58:*
93: em_esthum59:*
94: em_esthum60:*
95: em_esthum61:*
96: em_esthum62:*
97: em_esthum63:*
98: em_esthum64:*
99: em_esthum65:*
100: em_esthum66:*
101: em_esthum67:*
102: em_esthum68:*
103: em_esthum69:*
104: em_esthum70:*
105: em_esthum71:*
106: em_esthum72:*
107: em_esthum73:*
108: em_esthum74:*
109: em_esthum75:*
110: em_esthum76:*
111: em_esthum77:*
112: em_esthum78:*
113: em_esthum79:*
114: em_esthum80:*
115: em_esthum81:*
116: em_esthum82:*

117: gb_est48:*
 118: gb_est49:*
 119: gb_est50:*
 120: gb_est51:*
 121: gb_est52:*
 122: gb_est53:*
 123: gb_est54:*
 124: gb_est55:*
 125: gb_est56:*
 126: gb_est57:*
 127: gb_est58:*
 128: gb_est59:*
 129: gb_est60:*
 130: gb_est61:*
 131: gb_est62:*
 132: gb_est63:*
 133: gb_est64:*
 134: gb_est65:*
 135: gb_est66:*
 136: gb_est67:*
 137: gb_est68:*
 138: gb_est69:*
 139: gb_est70:*
 140: gb_est71:*
 141: gb_est72:*
 142: gb_est73:*
 143: gb_est74:*
 144: gb_est75:*
 145: gb_est76:*
 146: gb_est77:*
 147: gb_est78:*
 148: gb_est79:*
 149: gb_est80:*
 150: gb_est81:*
 151: gb_est82:*
 152: gb_est83:*
 153: gb_est84:*
 154: gb_est85:*
 155: gb_est86:*
 156: gb_est87:*
 157: gb_est88:*
 158: gb_est89:*
 159: gb_est90:*
 160: gb_est91:*
 161: gb_est92:*
 162: gb_est93:*
 163: gb_est94:*
 164: gb_est95:*
 165: gb_est96:*
 166: gb_est97:*
 167: gb_est98:*
 168: gb_est99:*
 169: gb_est100:*
 170: gb_est101:*
 171: gb_est102:*
 172: gb_est103:*
 173: gb_est104:*
 174: gb_est105:*
 175: gb_est106:*
 176: gb_est107:*
 177: gb_est108:*
 178: gb_est109:*
 179: gb_est110:*
 180: gb_est111:*
 181: gb_est112:*
 182: gb_est113:*
 183: gb_est114:*
 184: gb_est115:*
 185: gb_est116:*
 186: gb_est117:*
 187: gb_est118:*
 188: gb_est119:*
 189: gb_est120:*

190: gb_est121:*
 191: gb_est122:*
 192: gb_est123:*
 193: gb_est124:*
 194: gb_est125:*
 195: gb_est126:*
 196: gb_est127:*
 197: gb_est128:*
 198: gb_est129:*
 199: gb_est130:*
 200: gb_est131:*
 201: gb_est132:*
 202: gb_est133:*
 203: gb_est134:*
 204: gb_est135:*
 205: gb_est136:*
 206: gb_est137:*
 207: gb_est138:*
 208: gb_est139:*
 209: gb_est140:*
 210: gb_est141:*
 211: gb_est142:*
 212: gb_est143:*
 213: gb_est144:*
 214: gb_est145:*
 215: gb_est146:*
 216: gb_est147:*
 217: gb_est148:*
 218: gb_est149:*
 219: gb_est150:*
 220: gb_est151:*
 221: gb_est152:*
 222: gb_est153:*
 223: gb_est154:*
 224: gb_est155:*
 225: gb_est156:*
 226: gb_est157:*
 227: gb_est158:*
 228: gb_est159:*
 229: gb_est160:*
 230: gb_est161:*
 231: gb_est162:*
 232: gb_est163:*
 233: gb_est164:*
 234: gb_est165:*
 235: gb_est166:*
 236: gb_est167:*
 237: gb_est168:*
 238: gb_est169:*
 239: gb_est170:*
 240: gb_est171:*
 241: gb_est172:*
 242: gb_est173:*
 243: gb_est174:*
 244: gb_est175:*
 245: gb_est176:*
 246: gb_est177:*
 247: gb_est178:*
 248: gb_est179:*
 249: gb_est180:*
 250: gb_est181:*
 251: gb_est182:*
 252: gb_est183:*
 253: gb_est184:*
 254: gb_est185:*
 255: gb_est186:*
 256: gb_est187:*
 257: gb_est188:*
 258: gb_est189:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Thu Nov 8 09:02:33 2001

us-09-538-396-1.rst

Page 4

392 GCCCAACGCTGTCGCAAGACACATCATCGAGCTGCTGACCTTTCT 441
51 CysThrGluLeuProPheAsnSerArgSerGlyHisThrPheValHis 67
442 TGCAACGGCGAGCTGCCCCCACTCCGCTGCGCACACCTTCGTCA 491
67 BASPProLysValAlaGluGluThrGluThrGlyGluHisLeu 84
492 CGACCCCAAGGTAGCTGCGGAGCGAAGAAACAAAGACAAATTAGTTGC 541
84 rPheLysThrAlaAlaGlyLysAspValValCysIleArgSerPheGln 100
542 GGTTHAGACTGCAAGAGAAAGATGGTGTGTCATCCGCTCTTCAG 591
101 LeuThrGlnLysAlaSerLysMetGluPheLysAlaIleGluSerValle 117
592 CTTCACCAAAAGCATCAAAAGATGAGATTGAGCAATTGAAACGCTCT 641
117 uGlnThrIleAsnProHisThrGlyGluLysValCysLeuSerIleArg 134
642 CCAGACTTAATATCCACACAGGAGGAGAAAGTCTGCTCAGCTACAGAT 691
134 ySAlaAspMetAspArgGluIleProAlaLeuMetGlyValSerLysAla 150
692 GTGCGACATGATGATGAGAGATTCCTGCTTAATGGTGTTCAGAGCC 741
151 ValLeuGluAsnValIlePheValHisGlnAspGluSerAsnTrpProLe 167
742 GTACGGAGAAATGTTATTTGTTCCACCAAGATGATCAATGGCCATT 791
167 uGlnAspProSerThrLeuLysLysLysPheAspAspIlePheSerAla 184
792 GCAGAGACCCGTCACACTTAAGAAAGATTCAGATGACATCTTCTCTGCA 841
184 hrArgTyrThrLysAlaLeuGluValIleLysLysLeuHisLysAspGln 200
842 CACGCTTAACGAAAGCTTTGAACTCATAAAGAACTTCACAGAGATCA 891
201 MetGlnLysIleLysThrPheArgLeuLysLeuGluAsnLeuGlnThrVa 217
892 ATGCAAGAGATCAAGACTTTAGGTTAAAGCTGAGAACCTTCAGACTGT 941
217 LysAspGlnAlaHisLysLeuArgLysAsnIleAlaGlnAspGlnGlu 234
942 AAAGACCAAGACATTAAGCTGCGGAAATATTGCTCAAGATCAAGAA 991
234 ySerAspAlaSerLysSerGlnMetGluGlnLeuLysGluLysIleCys 250
992 AGTCAGATGCTCAAAATCTCAGATGAGCAACTGAGAAAGAAAGATCTGT 1041
251 GlyThrGluArgGluIleLeuGlnMetGluThrSerLeuAspGluLeuArg 267
1042 GGTACCGAGAGAGAAATCTCTCAATGAGAAACAAAGTTGATGAACCTAG 1091
267 ArgLeuGlnGlnGlnIleAspIleLysAlaThrGluArgSerThrLeu 284
1092 AAGACTTCAGGACAAATGACATCAAGCAACAGAGAGAAATGATTAAC 1141
284 euThrGlnGlnHisGluLysLeuAlaIleAspSerGluGluAsnGluAsp 300
1142 TTACGACACAGATGAAAGCTGCTGCACTTCTGAGGAAAGAAATGAT 1191
301 ThrAspGluGluLeuMetGluTrpGlnThrLysPheGluGluArgIleAla 317
1192 ACCGATGAGAACTAATGAAATGCAAAACAAATTTGAAGAAAGATATGC 1241
317 AleuLeuGluThrLysIleSerLysLeuValArgAspMetAspAspGlu 334
1242 GTTACTAGAAACAAATAACAGTAACTGTGAAGATATGATGTGAGAG 1291
334 lAspTyrSerSerValLeuSerLysGlnAsnSerGluLeuThrHisGlu 350
1292 CATCTTAATAGCTCGTCTGTCCAAACAAATTTGAAATTAAACATGAA 1341
351 IleGlyLysLeuGlnAlaGluAlaAspAlaHisLeuThrMetLysHisGlu 367
1342 ATTGAAAGCTCCAGAGAGAGCTGATGCTCAGCTGATCAAGAGATGA 1391
367 uArgAspSerAspIleLysAsnIleCysThrLysHisAsnLeuGlyPro 384
1392 ACGAGACTCAGACATTAATAATATATGACATTAACATATCTTGCGCCG 1441
384 aProGluHisProPheThrAsnAspValAlaMetAsnLeuThrAsnArg 400
1442 TTCCTGAACATCCCTTAACGAATGATGCTGATCAACCTTCAACAGG 1491
401 IleLysAlaArgLeuSerSerLeuGluAsnAspLeuLeuAspLysLys 417
1492 ATTAAGGAGACATCAAGCTTGAGATGATTTGCTGGATTAAGAGAA 1541
417 sSerAsnGluAspGlnLeuAspValLeuTrpLysHisTyrLeuLysIle 434
1542 ATCCAAATCAAGATCAGTATGATGTTTGTGGAACACATCTTAAATTA 1591
434 snAlaArgTyrSerGluValAspGlyGlnIleGlnSerLysIleGluSer 450
1592 ATGCTCGCTACTCCGAAGTTGATGTCAGATACAACTAAGATTGAATCC 1641
451 MetSerGlyIleLeuArgArgArgLysAspLysGluLysGluArgAspAl 467
1642 ATGTCAGGCATTTTAAGCGGAGAAAGATTAAGAGAAAGAACCGGAGC 1691
467 aAlaGluValGluLeuSerLysPheAsnLeuSerArgIleAspGluArg 484
1692 TGCAAGAGTGAAGCTTTCAAAATTAATCTATCCCGTATCCATGAGAGG 1741
484 LuArgHisMetGlnIleGluValGluArgLysThrLeuAlaLeuGlyGlu 500
1742 AGAGACATATCAAAATGAAGTCAGAGAGAGACACTGCGCTTCGAGAA 1791
501 ArgAspTyrAspSerIleIleSerGlnLysArgThrGluValTyrSerLe 517
1792 AGAGACTATGATTCAAATTAATTAAGTCAGAGAGACAGATATTAATAG 1841
517 uGluGlnLysIleLysValLeuLeuArgGluLysAspIleIleAsnArg 534
1842 GGAACAGAAATTAAGTGCCTTCGCGGAGAGAAATTAATTAATGA 1891
534 snAlaAspGluArgValLysLeuGlyLeuLysLysAspAlaLeuGluSer 550
1892 ATGCTGATGAAGAGATAAAGCTGGTTGAGAGAGATGATGGAAGC 1941
551 SerLysAspLysLeuAsnGlnIleValAsnGluHisLysAspLysIleLys 567
1942 AGCAAGGACAGCTCAATGAGATGATTAATGAGCTAAGATTAATGA 1991
567 sLysValLeuArgGluArgAsnProPheGluLysAspMetLysLysGlu 584
1992 AAAGTACTTGGGGGAGGAAATCTTTGACAGAGATTAAGAAAGGAGA 2041
584 leAsnGlnAlaPheTrpProValAspLysGluTyrAsnGluLeuArgSer 600
2042 TCAATCAAGCCCTTTGGCTGGGACAGAGATACAAATGAGTTAAGTCA 2091
601 LysSerGlnGluAlaGluGlnGluLeuLysPheThrGlnSerLysValThr 617
2092 AAATCCCAAGAGAGAGAGAGAGCTTAATTAATTTACTCAGAGAGAAATG 2141
617 rAspAlaArgGluGlnLeuThrLysLeuArgArgAspMetAspAlaLys 634
2142 TGATGCTAGAGACATTAAGCAAACTTCGAAGAGATATGATGCAAAA 2191
634 rArgPheLeuAspSerLysLeuGlnSerIleLeuGlnIleSerAlaAsn 650
2192 GAAGATTCCTCGAGCTGAAACTCAATCTAATTTACAGATATCTCTAAT 2241

651 ValAspMetPheProLysValLeuGlnAspAlaMetAsnLysArgAspG1 667
|||||
2242 GTTGACATGTTTCCAAAGTTCTACAGACGCATGACACAAAGAGATGA 2291
667 uGlnLysArgLeuGlnAspPheAlaAsnGlyMetArgGluMetLeuAlaP 684
|||||
2292 ACAGAAAAGATTGAGAAATTTCCAAATGGAATGCGGAAATGCTTGAC 2341
684 roPheGlnHisLeuAlaArgLysAsnHisValCysProCysGlyArg 700
|||||
2342 CTTTGAACATTTGGCTCGAAGAAATCATGTATGCCATGCTGTGAACGT 2391
701 AlaPheThrProAspGluGluAspGluPheValLysGlnArgMetG1 717
|||||
2392 GCTTACACCTGATGAGGAGGTAGTCTGTGAACAAAGATGCA 2441
717 nAsnSerSerThrAlaGluArgSerLysAlaLeuAlaMetGluSerSer 734
|||||
2442 AACTCAAGTACTGACAGAGATCTAAAGCTCTGCAATGGAATCATCA 2491
734 snAlaGluAlaLeuPheGlnGlnLeuAspLysLeuArgThrIleTyrAsp 750
|||||
2492 ATGCTGAACCTCTTTTTCAGCAATGGATAAACTTGCACTATCTATGAT 2541
751 AlaTyrValLysLeuValGluGluThrIleProLeuAlaGluLysAsnLe 767
|||||
2542 GCTTATGTAAGCTGTTAGAGAAACCATACCTTGACAGAGAAAACCTT 2591
767 uAsnGlnHisLeuAlaAspGluSerGlnLysAlaGlnAlaPheAspAspL 784
|||||
2592 GAATCAACATTTGGCGATGAAAGTCAAGAGGCGCAGCATTTGATGATC 2641
784 euLeuGlyAlaLeuAlaHisValGlnMetAspArgAspAlaValGluAla 800
|||||
2642 TTTTGGGTCTTCCCATGTTCAATGACAGAGGATGACAGGAGAACCC 2691
801 LeuLeuGlnProThrAspThrIleAspArgHisValHisGluIleGlnG1 817
|||||
2692 TTATTACAAACCCATGATCTATGACAGCATGTACATGAATTTACACA 2741
817 nLeuValLysGluValGluAspLeuGlnLysArgAlaLeuAspSerSerGly 834
|||||
2742 GCTAGCAAGAAGTAGAAGATCTTGAATATGCACTTGATTTGATGGCC 2791
834 rGcGlyValLysSerLeuGlnGluIleGlnLeuGlnLeuAsnPheLeuGln 850
|||||
2792 GAGGTGTCAAGCTTTGGAGAAATTCACACTGACCTGAACCTTTCGCAG 2841
851 ArgThrArgAspThrLeuIleValGluValAspAspLeuArgAspGlnH 867
|||||
2842 AGAACCAAGGCACATGTATGTCCAGAGTGAATCTTAGAGATCAACA 2891
867 sArgMetLeuAsnGluAspMetSerSerAlaGlnValArgTyrHisAsnA 884
|||||
2892 TGAATATGCTAAATGAAGATATGTCAGAGTCTCAGTGAGATGGCACAAG 2941
884 lArgGlnGluLysValLysAlaSerSerIleLeuGlnArgPheGlnLys 900
|||||
2942 CTCGGGAGAGAAAGTGAACCTTCTAGCATTTGGAAAGATTCACAAA 2991
901 SerGlnGluGluLeuValLeuLeuAlaGlnGluLysGlnGlnLeuIleVal 917
|||||
2992 TCTGAAGAGAAATGTGCTTCTAGCTGAGAAAAGAACCAACTGATTTGT 3041
917 lGlnLysLysLeuLeuGlnGluGluSerLeuAspProLeuSerLysGlnLysG 934
|||||
3042 AGAAAAGAGCTTTAGAAAGATCTCTTGATCATTTGTCCAAAGAGAAAG 3091
934 lUserLeuGlnGlnLysArgAsnAlaLeuLysGlnLysLeuAspGlnGlu 950
|||||
3092 AGAGCTTTGTGCAGAGATATATGCTTTGAACAAAAGCTGATGAAGAG 3141
951 TyrHisGlnLeuAlaGluArgLysArgGluPheGlnGlnGluLeuAspAl 967

|||||
3142 TATCATCAGCTTCAGAAAAGAAAAGGAGCTTCACACAGAACTTATGCG 3191
967 AluGlnLysArgLeuAsnMetLysIleLysGlyTyrLeuAspSerLysAla 984
|||||
3192 TCTTGAAGAGCTTAATATGAAATTAAGAGGTACTTGATTCACAGAAA 3241
984 snGlnLysLeuLysGlnLeuGlnGlyArgHisValLeuCysHisSerGln 1000
|||||
3242 ACCAAAAGCTTAAGGAATTCAGAGGAGGAGCATGTTCTTGCCATTCAG 3291
1001 LeuGlnSerCysMetAlaLysGlnGlnArgIleSerAlaGluLeuAsnLys 1017
|||||
3292 TTTCAGAGTTGCTATGCCAAGAACAGCAAGAAATATCAGTGAATTAACA 3341
1017 sSerLysGlnLeuLeuGlnGlnLysGlnGlnLysArgAsnIleAspAla 1034
|||||
3342 GAGCAAGAAGACTACTGACAGGCGCAGGCGCAGTTGAAGAAACATGAG 3391
1034 sPasnLeuLysTyrArgLysThrLysAlaAspValGluGlnLeuThrArg 1050
|||||
3392 ACATCTCAAGTACAGAAAACAAAGCTGATGTGGAACACTTACTCGT 3441
1051 AspIleGlnSerLeuGlnGluArgLeuLeuSerIleGlySerLeuSerAl 1067
|||||
3442 GATATGAATCACTTGAAGAAAGCTGCTTTCATATGATGCTGTCTGC 3491
1067 AlIleGlnAlaAspLeuLysArgHisSerGlnGlnLysGlnLysArgLeuAsn 1084
|||||
3492 TATAGAGAGCTGATCTGAAGGCGCATTTCTCAAGAAAAGAGGCTTAAT 3541
1084 eArgLysPheAsnArgTyrGlnGlyThrLeuSerValTyrGlnSerAsnIle 1100
|||||
3542 CAGAAATTTAAACAGGTGCAAGGAACACTTCTGTTATCAAGATATAT 3591
1101 SerLysHisLysGlnGlnLysLysLeuSerGlnTyrLysAspIleGlnLys 1117
|||||
3592 TCAGACCAACAAACAGAGCTTAACTGTACACATACAGAAATATGAGAA 3641
1117 sArgTyrThrAsnGlnPheLeuGlnLeuLysThrThrGluMetAlaAsnL 1134
|||||
3642 GCGATATACTAATCAATTTCTCCAGCTTAAGCAACTGAAAGGCAACA 3691
1134 yAspLeuAspArgTyrTyrThrAlaLeuAspLysAlaLeuMetArgPhe 1150
|||||
3692 AGGACTGTGACAGATATATATCTGCTTGTAGACAAGCTCTTATGCGGCTC 3741
1151 HisSerMetLysMetGlnGluLysLysIleIleLysGlnLeuThrGln 1167
|||||
3742 CACAGCATGAAGATGAGGAGATTAATAATTAATCAAGAACTGTGCA 3791
1167 nGlnThrTyrArgGlyGlnAspIleAspTyrIleSerIleAsnSerAspS 1184
|||||
3792 ACGAGATACAGAGCGCAGATATTTGATTAACATTAACATTAATTTGAT 3841
1184 eArgLysAlaAlaLysThrArgSerTyrSerTyrArgValAlaMetGlnThr 1200
|||||
3842 CTGAGGCTGCTGCACTCGATCATACAGCTACCGGCTTGTATGCAACT 3891
1201 GlyAspAlaGluLeuGlnMetArgGlyArgCysSerAlaGlyGlnLysVal 1217
|||||
3892 GGTGATGTGAGCTGGAATGCGAGGCGCTGACAGCTGCGTCAAGAGT 3941
1217 lLeuAlaSerLeuIleIleArgLeuAlaLeuAlaGlnLysThrPheCysLeu 1234
|||||
3942 TCTGTGTTCTTATATACAGCTACAGCTTCCGAAACTTCTCTGCTGA 3991
1234 snCysGlyIleLeuAlaLeuAspGluProThrThrAsnLeuAspGlyPro 1250
|||||
3992 ACTGCGGTATATTTGCTTGTGATGAGCCAACTAACATCTAGATGGCCA 4041
1251 AsnAlaGlnLysLeuAlaAlaLeuLeuArgIleMetGlnAlaArgLys 1267
|||||

1180 AATAATGAACCTTGACATGAATTAAGCCTTGATAGCCGAAGAGC 1229
282 htleuethrInglInHISgluLysleuAlaLeuSerGluGluAsn 298
1230 AATGGAGAAAGATTAATAGTGAAGCTGAAGAGAAATGAAAAAGTTT 1279
299 GluAspThrAspGluGluMetGluTrpGluThrLysPheGluGluAsp 315
1280 CAAGGAGCTGATGACCACTAAATGACTTATATACAAATCAGAGAAC 1329
315 gtleAlaLeuLeuGluThrLysLysSerLysLeuValArgAspMetAsp 332
1330 AGTAAGGAGAAAGAAAGAAATGGTAGAGCTGATCGTGAAGCTGAAA 1379
332 spcLysAspTyrSerSerValLeuSerLysGlnAsnSerGluLeuThr 348
1380 AACTAAATAAAGAACTAGGCTTCTCAATCAGAAAAATCAGAACTGCTT 1429
349 HIsGluLysGluLysLeuGlnAlaGluAlaAspAlaHIsLeuThrMetLys 365
1430 GTTGAACAGGCTGCTACAGCTGCAAGAGATCCCATCAAGACATAT 1479
365 sHIsGluArgAspSerAspLysLysAsnLysCysThrLysHIsAsnLeuG 382
1480 CCGAGCTAGAGATTCATTAATTCAGTCTTGCGCAACAGACTGAATGG 1529
382 LysProValProGluHIsProPheThrAsnAspValAlaMetAsnLeuThr 398
1530 ATGGCTTTGAGCGTGCAGCATCTCAGTGAAGACAGATTAATAATTTT 1579
399 AsnArgLysLysLysLysLysLysLysLysLysLysLysLysLysLys 408
1580 AAACCTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1629
408 uGluAsnAspLeuLeuAspLysLysLysSerAsnGluAspGlnLeuAsp 425
1630 GATGAATGACTTTCAGAAAAAGAGACTGGAACAAACAAACAGATGAT 1678
425 alleuThrLysHIsLysLysLysLysLysLysLysLysLysLysLys 441
1679GAGTAAAGATTAAGAAACTGAGACTGGA 1708
442 GlyGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLys 458
1709 AGAATTAATGAGTAAATCAGAAATCTAAGT.....AA 1743
458 GluAspLysGluLysGluArgAspAlaAlaGluValGluLeuSerLys 475
1744 GAAGCAGATGAGCTGAAAAATGTGAAGTGAATTAACAGCACTGGAAG 1793
475 heAsnLeuSerArgLysAspGluArgGluArgHIsMetGlnLysGluVal 491
1794 GATCTTCAGAGATTTCTGAGCTGAGCAGAGCTC...ATAAAGCT 1840
492 GluArgLysThrLeuAlaLeuGluGluArgAspTyrAspSerLysLeu 508
1841 GAACGAGAG.....TTAAGCAAGCTGAGAAAAACAGCAATGTGA 1881
508 rgLysArgThrGluValTyrSerLeuGlnLys..... 520
1882 AACCTTAATAAATGAAGTAAATAGTCTCCAAATGAAAAAGCAGACTAG 1931
521LysValLeuLeuArgGluLysAspLysLysLysLysLys 534
1932 ACAGAGACCTGCGTAACTGACAGAGATGAGCAGCATTAACCAT 1981
535 AlaAspGluArgValLysLeuGluLysLysLysAspAlaLeuGlu 550
1982 ACACACACACGTACCCAAATGAGATGCTGACCAAGACAAAGCTGCAA 2031
550 rSerLysAspLysLeuAsnGluLysValLeuGlnHIsLysLysLysLys 567
2032 AGAT...GAACAAATCAGAAAAATTAATCTAGCACAGTGAATTA 2078
567 YsLysValLeuArgGlyArgAsnProPheGluLysAspMetLysGlu 583
2079 CTTCACTGTG...GGATATTTTCCACAAAAAGCGTTGAAGAC... 2122
584 IleAsnGlnAlaPheTrpProValAspLysGluTyrAsnGlnLeuArg 600
2123TGG.....CTACATAG 2133
600 rLysSerGlnGluAlaGluGlnLysLysPheThrGlnSerLysVal 617
2134 TAAATCAAAAGAAATTAATCAG..... 2155
617 htrAspAlaArgGluGlnLeuThrLysLeuArgArgAspMet..... 630
2156ACCAGGACAGCTTGGCCAAATGTGAACAGAGAACTAGCTTATCT 2200
631 AspAlaLysArgArgPheLeuAspSerLysLeu...GlnSerLysLeu 646
2201 GACAGCAATTAATAATCATTAATTAATGAACCTAAAAAGAAAGAGCA 2250
646 nIleSerAlaAsnValAspMetPheProLysValLeu.....GlnAsp 661
2251 GTTGTCAGATTACGAAAGCAAGCTGTTGATGTTGTGTAAGCCAGAT 2300
661 IametaAsnLysArgAspGluGlnLysArgLeuGlnAsnPheAlaAsnGly 677
2301 TTGAAGCTGATTAGACAGCTTAAAGAGAAATTAATAATCATCAAAA 2350
678 MetArgGluMetLeuAlaPro.....PheGln 686
2351 CAGGAGCCATGCTGCTGAGGACACAGAGCTTACTCCAGTTCATTAC 2400
686 uHIsLeuAlaArgLysAsnHIs...ValCysProCysGluArgAla 702
2401 TCAGTACAGAGAGAAACAGATGATGCTGCGGTTTCAGAGAGTTT 2450
702 he...ThrProAspGluGluAspGluPheValLysLysGlnArgMetGln 717
2451 TTCAGACAGAGGCTGATTCAAGAAAGTCACTGATGATTTGCGTCTAA 2500
718 AsnSerSerThrAlaGluArgSerLysAlaLeuAlaMetGlnSerSer 724
2501 CTGCGACTTCTCCAGTAACTCAAGTCAACAGAAATCAGAGCTAAAAA 2550
724 nAlaGluAlaLeuPheGlnLysAspLysLysLysLysLysLysLys 751
2551 AAAGAAAAAGCGCGTGAATTAATGCTGGAGCTTGCCCATGAGGCAA 2600
751 LysValLysLeuValGluGluThrLysPro...LeuAlaGluLys... 765
2601 GCATTAATTTGATTTGAGAGAGAAATACCAAAATTAAGAAACAACTG 2650
766 ...AsnLeuAsnGlnHIsLeuAlaAspGluSerGlnLysLysAlaPhe 781
2651 CAGATGTCATTAAGACATACAGCGCTTAAGAGACATAGAGAAACA 2700
781 eAspAspLeuLeuGly..... 786
2701 AGAAACACTTGGGTACATAATGCTGAAAGAAAGTCCCAAGTAT 2750
787ValLeuAlaHIsValGlnMetAspArgAspAla 797
2751 GCCTGACAGATGTTACAAATTAATGAGAGCTTCCAGATGCAACTAAAGAT 2800
798 ValGlu.....AlaLeuGlnProThrAsp 807
2801 GTTGAAGAAAAATTCGACACAGAGCTTAACCTCAAGAAATAGAC... 2848
807 rIleAspArgHIsValHIsGluLysGlnLysGlnValGluAla 824
2849 .TAGATGCAAGCTGCCAACAGTCAACAG.....GAAAAACAAG 2888

```

824 splenuluytraleuaspsercylargylvalylserleu 840
      ::::::::::::::::::::
2889 AGAAGACGACAGTATGACACA.....GTTCTAGTAAT 2926
      ::::::::::::::::::::
841 glutileuglualeuasnpheluglnargthargapthleu 857
      ::::::::::::::::::::
2927 GAATTGATCTGAGCTT..... 2944
      ::::::::::::::::::::
857 evalgluvalasprleuargapghisargmetleuasnqluapm 874
      ::::::::::::::::::::
2945 .....ATACAGACGACGAGAACAGATTCACATCTAA 2978
      ::::::::::::::::::::
874 etseralaglnvalargthargasnlaargglugluvalyls 890
      ::::::::::::::::::::
2979 AAAGTACAAAC.....AATGAGCTAAATCTGAGAACTTCAG 3016
      ::::::::::::::::::::
891 Alaserleleugluargphelugllyserglugluvalle 907
      ::::::::::::::::::::
3017 ATATCCACTATTTGCAACGTCGACGAACTGAGGAGCACTGTGGA 3066
      ::::::::::::::::::::
907 uleuaglucgluglualeuilevalglulylserleu 924
      ::::::::::::::::::::
3067 ATATCCACTGAGTCTGCTTTGTACAGAGATTAAGATCTAAAG 3116
      ::::::::::::::::::::
924 luserleuaproleu.....serlysgluys 933
      ::::::::::::::::::::
3117 AGCAGGTAGGCCCTTGGAAACAACTGGAAAGTCCACCAAGAAA 3166
      ::::::::::::::::::::
934 gluserleuugluargthargasnleuargllyleuaspqlugl 950
      ::::::::::::::::::::
3167 GAGATTAATCAACAAAAAATATACAGCAAAATAGCACAGATTA 3216
      ::::::::::::::::::::
950 uyrhislgluvalgluargllyarggluphelnuglualeuap 967
      ::::::::::::::::::::
3217 ACTGATATATTAAGGAGGTTAAATAATTCATGCTATATG... 3262
      ::::::::::::::::::::
967 lalenglargleuasnmetlyslileysgluytrleuaspserlyys 983
      ::::::::::::::::::::
3263 .....AAAGCATGTGAATTTATTCATCAAGATGGGAAA 3295
      ::::::::::::::::::::
984 asnglu.....lyslleuylgluvalgluargllyarghlsvalleucysn 998
      ::::::::::::::::::::
3296 GACGCTATTAAGACCAAAAGAACTGACTTAATAAGTAAAT... 3340
      ::::::::::::::::::::
998 ssergluenglusercysmetalaalysglnglnarglleserlaglul 1015
      ::::::::::::::::::::
3341 .GCTCACTAAGTGAATGCGAGAAACACAAAGAAATTAATGAAATGA 3389
      ::::::::::::::::::::
1015 euasnlyserlysglualeuenglngllynglgluvalylsargasn 1031
      ::::::::::::::::::::
3390 TGAGACTCATGACAGACAGATATGTACACAGAGATACAAAGAAAGTGG 3439
      ::::::::::::::::::::
1032 llearpasphenleuyltyrargthargasnlaasprvalgluglnle 1048
      ::::::::::::::::::::
3440 CTACAGATTAACCTTACTTAAGAAAAGAAATGAGAACTAAAGAAAGT 3489
      ::::::::::::::::::::
1048 uthrargaprllegluserleuugluargleuuserllegllyserl 1065
      ::::::::::::::::::::
3490 TGA.....GAAGAAAGAAACAACTTTGAAGGAAATGGGCTCAA 3530
      ::::::::::::::::::::
1065 euseralalegluvalasprleuysarghlserserlglulys..... 1079
      ::::::::::::::::::::
3531 TCCAGGCTTTGCAAAATGAAAAGTGAACATCAAGAACTGGAGAGAGACATA 3580
      ::::::::::::::::::::
1080 gluargleuasnserglupheasnargtrgngllythrleuservalty 1096
      ::::::::::::::::::::
3581 GACATATATAAAGAAATCATATTTGGCATTAAGGCGACAGAAAGGTGA 3630
      ::::::::::::::::::::
1096 rglinseranlleserlyvnhlsylsnglgluvalylserglntyl 1113
      ::::::::::::::::::::
3631 TGAAGAGAAATTAATTCATTTAAGAAAGAACTGAGAACCAATATTC 3680
      ::::::::::::::::::::
1113 yasprleuylusargtyrthrargnglnphelenglleuylserthr 1129

```

```

      ::::::::::::::::::::
3681 GCGATCGTGAAGAAAGATATAGAGAAATGATGATTGTATGAGGACACACA 3730
      ::::::::::::::::::::
1130 Glumetalasnllyasprleuapargtyrtyrthrallaleuaplysal 1146
      ::::::::::::::::::::
3731 GAACCTTGAAACAGAGATCTGATATTATTAAGACTCTTGACCAAGC 3780
      ::::::::::::::::::::
1146 aleueterargphetissermetlysmetglugluileasnlysllelel 1163
      ::::::::::::::::::::
3781 AATATATGAATTTTCACAGTATGAAATGAAATGAAATCAATTAATTTATC 3830
      ::::::::::::::::::::
1163 ysgluuerytrpgnglnthrtyrarggllynglnasprlleasprtlleser 1179
      ::::::::::::::::::::
3831 GTGACCTGTGGCGAAGTACCTATCTGACCAAGATATTGATACATAGAA 3880
      ::::::::::::::::::::
1180 lleanserasp.....sergluylalagluythrargse 1191
      ::::::::::::::::::::
3881 ATACGGTCTGATGCCGATGAATAATGTATCAGCTCTGTAAAGCGGAA 3930
      ::::::::::::::::::::
1191 rtyrserlyrargvalvalmetglnthrlyaspalagluluglunleueta 1208
      ::::::::::::::::::::
3931 TTATTAACCTACGAGTGTGATGTGAGAGGAGACACAGCCTTGATATGC 3980
      ::::::::::::::::::::
1208 rgllyargcysseralaglyngllyvalleuajaserleuilelearg 1224
      ::::::::::::::::::::
3981 GAGACAGATGACGTGCTGACAAAGATATTAGCCTCATCATCTCGC 4030
      ::::::::::::::::::::
1225 leuajaleuajagluthrphecysleuasnysgllyleuajaleuajals 1241
      ::::::::::::::::::::
4031 CTGGCCCTGCTGAACGTTCTGCTCAACTGTGACATCATCTGCTGGA 4080
      ::::::::::::::::::::
1241 pgluprothrthrleuasnaspqluproasnagluuserleuajalaal 1258
      ::::::::::::::::::::
4081 TGAGCCAAACAAATCTTGACCGAGAAACATGAACTCTTGACACTG 4130
      ::::::::::::::::::::
1258 lalenuarglleuergluajlaargllysgllynglgluvalasnphelnleu 1274
      ::::::::::::::::::::
4131 CTCGGTGAATATATAAAAGTCCGACAGACGCTCACTCAGCTT 4180
      ::::::::::::::::::::
1275 llevallyethrhlaspqluargphelahlslleuilegllynglnargl 1291
      ::::::::::::::::::::
4181 CTGTAATCACTCATGATGAAGATTTGTGGCCTTTAGAGCCTTGA 4230
      ::::::::::::::::::::
1291 nleuajagluysrtyrtyrargvalserlyasprgluasnghlsersi 1308
      ::::::::::::::::::::
4231 ATATGTGAGAAATTTCTACAGATTAATAAGAACTCATCATGCTCTGAG 4280
      ::::::::::::::::::::
1308 leile 1309
      ::::::::::::::::::::
4281 AGATT 4285
      ::::::::::::::::::::
seq_name: /SDSL/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV59979
seq_documentation_block:
ID AAV59979 standard; cDNA to mRNA; 5893 BP.
XX
AC AAV59979;
XX
DT 25-NOV-1998 (first entry)
XX
DE Nucleic acid encoding a human homologue of yeast RAD50.
XX
KW Human homologue; yeast RAD50; Drosophila Septin-2; Acyl-CoA synthetase;
KW Immunomodulatory activity; Identification; activated T-cell; ds.
XX
OS Homo sapiens.
XX
FH key Location/Qualifiers
FT CDS 389..4325
FT /'tag= a
PN WO9838306-A1.
XX

```


425 aileutrllyhlsrlytleuylsleashnlahrgylsergluvlaap 441
1679 :gagatagagatgaataagaaactggacgtggca 1708
442 glylnileglnserlylelglsrmetserglylleuarlgar 458
1709 agaatatttgactttaaatcagaaattcctaagt aa 1743
458 glyasrlysglulysgluargapalaalaglualeuenserlyp 475
1744 gaagagatagactgaaaaaatggaagatgattcacacagtggag 1793
475 heasnluserlyrlyleasplulrglunarghismetglnleglval 491
1794 gatcttcagacaggtttcttaactgtgcacggagcgc .. atpaaagct 1840
492 gluatgylsrthleualleuglyglunrgrapyraasrlellese 508
1841 gaaactgag tttagcagaagctgagaaaaacgcaattgaga 1881
508 rglulysarqthrgluvaltryserleuglnlys 520
1882 aaccttaaaatggaatataagactccaataagaaaagacacttag 1931
521 lleyvalleuuarqglulysapllelleaargasn 534
1932 acaggaacctggctgaactgtaccgagatggagcaggttmaaccatcat 1981
535 alaaarglualrgvallyseugly .. leuylsyrasapalaleuglse 550
1982 acacacacacgtaccctaaatggaatgctgaccaaagcaaaagctgaca 2031
550 rserlyasrlyleuaingluilevalasnglnhlsrlyasrlylel 567
2032 agat .. gaaacaatcacaaaaatttaattctagacacagtgatgaattaa 2078
567 yslvalleuarqglulargapnropheglnlyasrmetlyslvglu 583
2079 cctcaactgttg .. ggaatattttcccaacaaaacacacttggaagac .. 2122
584 lleasnglnlahpethrprovalasrlysglulryanglualeuarqse 600
2123 tgg ctacactag 2133
600 rlysserlnglualeaglnglnleuylsrphetrhnglnserlyalt 617
2134 taaattcaaaagaaatttaactag 2155
617 hraspalahrglucnleuthrlyleuarlgarqasrmet 630
2156 accagagacagacttgcctaaattgaacaaagaaactgcttcact 2200
631 asrpalalyargargrphelueusrsrlyseu .. glnsrilleugl 646
2201 gacacaaatpaaaaatcatatttaattatgacttataaaagaggaagaca 2250
646 nilesrlelaasvalasrmetpethrlyslaleu glnaspa 661
2251 gttgtccactgacgaagacaaactgtttgagttgtgtgtgacacagatt 2300
661 lamelanslyargarqglulnlyrlygleugluasnphelalaangly 677
2301 ttgaatgatgtttagacagcctttaaaagagaaattgaaaattcatcama 2350
678 metarglulmetleuabapro phegl 686
2351 caccgagaccactgtgctgctgagaccacacagattttactccacgttcattac 2400
686 uhlslleualalygluadanh .. valcyarocysyvgluatgalap 702
2401 tcaacttaacagcaaaacacactcatgttgcccgctgtgtgacaaagt 2450

702 he...thpRoasprbubusprgubhpheValysLysGlnArgMetGln 717
 2451 TTTCAGACAGAGCGCTGATTACAGAAAGTCATCAGATGTCGACTCTAA 2500
 718 AasSerSerThAlaGlnArgSerLysAlaLeuAlaMetGlnSerSer 734
 2501 CTGGCAGCTTGGTCCCGATTAACCTCAAGTCACACAGATTCAGAGCTAA 2550
 734 naLaGlnAlaLeuPheGlnGlnLeuAspLysLeuArgThrIleTyrAsp 751
 2551 AAGGAAAGCGCGCTGATGAATCTGGACTGTGCCCTGAGGCAAA 2600
 751 IatYValLysLeuValGlnIuThrIlePro...LeuAlaGlnLys... 765
 2601 GCATTAATGATTATTGAAGGAAAGAAATTCACGAATTAAGAAACAAC 2650
 766 ...AsnLeuAsnGlnHisLeuAlaAspLysSerGlnLysAlaGlnAlaP 781
 2651 CAGATGTCAATATAGACATACAGCGCCTTAAGAAAGCATAGAAAGAA 2700
 781 eAspAspLeuLeuGly... 786
 2701 AGAAACACTCTGGGTACATATGCTGGAAGAAAGAAAGTCCAAAGTAT 2750
 787 ...ValLeuAlaHisValGlnMetAspArgAspAla 797
 2751 GCCTGACAGATGTTCAAATTAATGAGAGGTTCCATGGAACCTTAAGAT 2800
 798 ValGln...AlaLeuLeuGlnProThrAspThr 807
 2801 GTTGAAGAAAGAAATTCACACACACAGCAGCTAAGCTACACAGAAATAGC 2848
 807 rIleAspArgHisValHisGlnIleGlnIleLeuValLysGlnValGln 824
 2849 ...TAGATGCAGCTGGTCCAAACAGTCAACAG...GAGAAACAG 2888
 824 sPLeuGlnTyrAlaLeuAspSerSerGlyIatYValLysSerLeuGln 840
 2889 AGAAACAGACAAAGTATGAGCACAA...GTTTCAGTAAGTAT 2926
 841 GlnIleGlnLeuGlnLeuAsnPheLeuGlnIatYThrArgThrLeuI 857
 2927 GAATTAATTCGTAAGCTT... 2944
 857 eValGlnValAspAspLeuArgAspGlnHisArgMetLeuAsnGlnAsp 874
 2945 ...ATACAGACACAGCAGACAGATCAATCAATCTAA 2978
 874 eIserSerAlaGlnValArgThrPHisAsnAlaArgGlnGlnLysValLys 890
 2979 AAGTACACAA...AATGACCTTAATATCTGAGAACTCTG 3016
 891 AlAspSerIleLeuGlnIatYArgPheGlnLysSerGlnGlnIleValLe 907
 3017 ATATCCACTAATTGGACAGCTGTCAGCAGTCGAGAGACACACTGTGGA 3066
 907 uLeuAlaGlnGlnLysGlnIleLeuValGlnLysLysLeuLeuGln 924
 3067 ATATTCACAGAGTTCACAGTCTTTGTACAGAGAGATAAAGATGCTTAAG 3116
 924 IuSerLeuAspProLeu...SerLysGlnLys 933
 3117 AGCAGTAGAACCCCTTTGGAAACAACTTGGAAGAGTTCCAGCAAAAA 3166
 934 GlnSerLeuLeuGlnIleLysIatYArgHisAlaLeuLysGlnLysLeuAspGlnI 950
 3167 GAAGAAATTAATCAACAAAAATTCACAGCCACAAATATGCGCAGAGATA 3216
 950 uTyrHisGlnLeuAlaGlnIatYArgLysArgIlePheGlnGlnIleuAsp 967
 3217 ACATGAATGATTAATTAAGACAGAGTTAAAAATATTCATGCTAATAG... 3266
 967 IatLeuGlnIatYLeuAsnMetLysIleLysGlnTyrIleuAspSerLysLys 983

```

3263 .....AAAGACATGAGATATATATTCAGATGGGAAA 3295
984 AsnGlu.....LysLeuYsgLuleuGlnGlyArgHisValLeuCySH 998
3296 GACGATATAGAGAGCAAAAGAACTGAATTAATTAAGTATA..... 3340
998 sserGlnLeuGlnSerCysMetAlaLysGlnGlnArgHisLeuAlaGlu 1015
3341 .CTCAACTAAGTGAATGCGAGAACACACAAAGAAATTAAGAAAGATA 3389
1015 euAsnLysSerLysGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGln 1031
3390 TGAGACTCTAGACAGATATGATACACAGATACAGAAAGAGTGG 3439
1032 TLeaspAspAsnLeuLysTyrArgLysThrLysAlaAspValGlnGln 1048
3440 CTACAAAGATACCTTACTTAAAGAAAGAAATGAGAACTAAAGAAAGT 3489
1048 uThrArgAspLysLeuSerLeuGlnGlnGlnGlnGlnGlnGlnGln 1065
3490 TGA.....GAAGAAAGAAAGAAAGAAAGAAAGAAAGTGTCAAA 3530
1065 euSerAlaLysGlnLysArgHisSerGlnGlnGlnGlnGlnGlnGln 1079
3531 TGCAGGTTTGCAGAAAGTGAACATCAGAGTGGAGAGAGACATA 3580
1080 GluArgLeuAsnSerGlnPheAsnArgTyrGlnGlnGlnGlnGlnGln 1096
3581 GACAAATATTAAGAAATCATTAATTTGCGATTAGGCGACAGAAAGTTA 3630
1096 rGlnSerAsnLysSerLysHisLysGlnGlnGlnGlnGlnGlnGln 1113
3631 TGAAGAGAAATTAATTCATTAAAGAAAGAAAGTTCGAGAACCAATTTTC 3680
1113 yAspLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1129
3681 GGGATGCTGAGAGAAAGTATAGAAATGATGTTTATGAGACAGAA 3730
1130 GlnMetAlaAsnLysAspLysAspArgTyrTyrThrAlaLeuAspLysAl 1146
3731 GAACCTGTGACAGATCTGATATTTATTAAGACTCTTGACCAACG 3780
1146 AleuMetArgPheHisSerMetLysMetGlnGlnGlnGlnGlnGlnGln 1163
3781 AATTAATGAATTCACAGATGATAAAGAAAGAAATTAATTAATTAATTC 3830
1163 yGlnLeuTyrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1179
3831 GTGACCTGTGGGAGATACCTATCTGTGACAGATATTAATTAATTAATTA 3880
1180 ILeasnSerAsp.....SerGlnGlnGlnGlnGlnGlnGlnGlnGln 1191
3881 ATACGCTGTGATGCGAGTGAATGATGATGATGATGATGATGATGATG 3930
1191 rTyrSerTyrArgValAlaMetGlnThrLysAspAlaGlnGlnGlnGln 1208
3931 TTATATACACACAGTGTGATCTGAAAGGAGACACAGCTTGATGATGTC 3980
1208 rGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1224
3981 GAGGACGATGCTGCTGACAAAGGATTTAGCTCTCATCTCATCTCC 4030
1225 LeuAlaLeuAlaGlnThrPheCysLeuAsnGlnGlnGlnGlnGlnGln 1241
4031 CTGGCCTGTGGGAAAGTGTGCTCAACCTGAGCATCATGCTGTGA 4080
1241 PGLuProThrThrAsnLeuAspGlyProAsnAlaGlnSerLeuAlaAla 1258
4081 TGAGCCACACAAATCTTGACCGAGAAACATGATCTTGTGACATG 4130
1258 lAluLeuArgLysMetGlnAlaArgLysGlnGlnGlnGlnGlnGlnGln 1274

```

```

4131 CTCTGTTGATATATTAATTAAGTGTGCTCAACAGAGCTTAAGTCAAGCTT 4180
1276 TLevalIleThrHisAspGlnArgPheAlaHisLeuIleGlnArgGln 1291
4181 CTGTATATCATCATGATGATGATTTTGTGACCTTTTGTGACCTTCTCA 4230
1291 nLeuAlaGlnLysTyrTyrArgValSerLysAspGlnuAsnGlnHisSer 1308
4231 ATATGTGAGAAATTTCTACGATTAATAAGAAATGATGATGATGATG 4280
1308 lIle 1309
4281 AGAT 4285

seq.name: /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AAA09326
seq.document_block:
ID AAA09326 standard; DNA; 6773 BP.
XX
AC AAA09326;
XX
DT 10-AUG-2000 (first entry)
XX
DE Human cancer associated antigen precursor DNA, clone NY-REN-53.
XX
KW renal cancer; cancer associated antigen precursor; diagnosis;
XX cytosolic; KIAA0336 mRNA tag; ss.
XX
OS Homo sapiens.
XX
PN WO200020587-A2.
XX
PD 13-APR-2000.
XX
PF 04-OCT-1999; 99WO-US22873.
XX
PR 05-OCT-1998; 98US-0166300.
XX
PR 05-OCT-1998; 98US-0166350.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Obata Y, Golt I, Turci O, Sahin U, Pfeundschn M, Scanlan MJ;
PI Stockert E, Chen Y, Old LJ, Jager E, Knuth A;
DR WPI: 2000-303774/26.
XX
PT Preventing, diagnosing and/or treating disorders associated with
PT abnormal expression of human cancer associated antigens
XX
PS Claim 57, Page 97-99; 121pp; English.
XX
XX AAA09321-45 were isolated by SEREX screening from a renal cancer
XX cell line 1973/10.4. Homology searching revealed that these clones
XX correspond to known genes. The present sequence has identity with the
XX KIAA0336 mRNA tag gene. The genes encode cancer associated antigen
XX precursors. These gene products are useful in methods for preventing,
XX diagnosing and/or treating disorders, especially cancer, associated with
XX abnormal expression of human cancer associated antigens. The method
XX comprises contacting a sample from a subject with an agent that
XX specifically binds to the nucleic acid molecule or expression product
XX (or fragment) complexed with a human leukocyte antigen (HLA) molecule
XX and determining the interaction between the agent and the nucleic acid
XX molecule or the expression product as a determination of the disorder.
XX
SQ Sequence 6773 BP; 2566 A; 1092 C; 1363 G; 1752 T; 0 other;

alignment_scores:
Quality: 478.50 Length: 1249
Ratio: 0.700 Gaps: 59
Percent Similarity: 54.764 Percent Identity: 21.057

alignment_block:

```

US-09-538-396-2 x AAA09326

Align seg 1/1 to: AAA09326 from: 1 to: 6773

```
91 LysAspValValCysIleArgSerPheGlnLeuThrGlnLysAlaSerLy 107
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
707 AAAGAGTGTGATGTC.....CAGATGAGACATCAGCTAAGAAACATGA 750
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
107 smetGluPheLysAlaIleGluSerValLeuGlnThrIle.....Asp 122
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
751 AGCAGAGTAATTAAGTGAACGAGCTAAAGAAACACTTAAACAAAT 800
122 roHsThGlyGlyLys...ValCysLeuSerTyrArgCys..... 134
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
801 GTGAGGCAAGTGAAGAAACATCCAGAAACATATGAAATGAGCTTGA 850
135 .....AlaAspMetAspArgGluIleProAl 143
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
851 AATTAAAGAAAGCCACCTCAATGCAACCAACCAACATCAGATATGTC 900
143 AluMetGlyValSerLysAlaValLeuGluAsnValIlePheValHisG 160
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
901 TATCTCTTGCAGAAATAATACATTGTGTGAACAAAGTAGTA..... 940
160 LnaAspGluSerAsnTyrProLeuGlnAspProSerThrLeuLysLys 176
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
941 ..AMGAAAGATCAACACCTTAGAAGAT.....ACCTTAAAGAA... 979
177 PheAspAspIlePheSerAlaThrArgTyrThrLysAlaLeuGluVal 193
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
980 .....CTTGAATCTCA 990
193 eluLysLysLeuHisLysAspGlnMetGlnIleLysThrPheArgLeu 210
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
991 ACACAGTATCTTAAAGAGTGAAGTATATGATATATCTTAAAGTTAA 1040
210 yslLeuGln...AsnLeuGlnThrValLysAspGlnAlaHisLysLeu 225
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1041 AACTGGAATGATGCTCAACATATTAAGAGATGATTTTTCATGAACG 1090
226 GluAsn.....IleAlaGlnAspGlnI 233
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1091 GAAAGCTTAGAGTTAAATTAATGAATATATCTAGCTAAGAAAGAA 1140
233 uLysSer.....AspAlaSerLysSerGlnMetGlnIleLysGlu 248
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1141 GGGCTGTGTAATGAAATAATTAATCTGACCTAGCAGGTTAAATAAC 1190
248 yslIleCysGlyThrGlnArgGluIleLeuGlnMetGlnThrSerLeu 264
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1191 AGTTTGTCTATCTGTAGAACAGATTAACAGAAAGTACAGAGCTTAA 1240
265 GluLeuArgArgLeuGlnGlnIleAspIleLysAlaThrGluArgSe 281
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1241 GAA.....CAACATCAAAAAAATAATATCAGAACTAAATAGAGAC 1278
281 rThrLeuLeuThrGlnGlnHisGlyLysLeuAla..... 292
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1279 ATTTTGTTCAGATTCAGAAAAAGAAAAATTAACATTAACTTTGAATAC 1328
293 ..AlaLeuSerGlnLysGlnAsnGluAspThrAspGlnLysLeuMetG 308
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1329 AGGGCTTAAGACAGCTGTAACCTTACAGCAAGAAAGCAACAAAGCA 1378
309 GlnThrLysPheGlu.....GluArgIleAlaLeuLeuGlnThrLy 322
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1379 AATTAAATTAATAGAGCTTAGAGAGATTAAGAAATTTTCAAAACAGA 1428
322 sIleSerLysLeuValArgAspMetAspAspGluAlaSerTyrSerSe 339
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1429 ACTGGGGGAATCTCTGCAAAATAATAGTCMAAG.....TTGGAATCA 1472
339 AluLeuSerLysGlnAsnSerGluLeuThrHisGln..... 350
```

```
1473 TGAAGCAACAGACAGCATCTGATGTT...CATGAATCGACAGACAGCTC 1519
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
351 .....IleGlyLysIle 354
1520 AGAAGCTCTTTTACTGAAAAAGATGCCCTTCTGAAACCTGGAATGCCCT 1569
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
354 uGlnAlaGlnAlaAspAlaHisLeuThrMetLysHis.....Glu 368
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1570 CCAGGAGAAAAATGAAGATTAATCTATCTCAACAGATTTGGTACCAGAAC 1619
368 rgaSerAspIleLysAsnIleCysThrLysHisAsnLeuGlyProval 384
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1620 TTGAATATACATTAAGAACCTTCAGAAAAAGAAATGAGATTA..... 1660
385 ProGluHisProPheThrAsnAspValAlaMetAsnLeuThrAsnArg 400
1661 .....TACTTACTTAGTCTCAGTCAAGCA 1686
401 .....IleLysAlaArgLeuSerSerLeu.....G 409
1687 TACCATGTTAAAGAAATTAAGAGAAAGATTAATCTTCTAGAGAA 1736
409 LnaAspLeuLeuAspLysLysLysSerAsnGluAspGlnLeuAspVal 425
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1737 AAGATGATTTATTAATTAATTAAGTAAATCTCCATGAAGAAATGATAT 1786
426 LeuThrLysHis.....TyrLeuLysIleAsnAl 435
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1787 TTCCATTAAGAAATGTGAAGGAAAGAAAGATTTGATTTCTAGACTTGG 1836
435 AsnGlySerGluVal.....AspGlyGlnIleGlnSerLysIleG 449
1837 GAAAGTACAGCAACAAATCCAGTACACAGTCACTAGACAAACAAAGTAA 1886
449 LysSerMetSerGlyIleLeuArgArgArgLysAspLysGluLysGlu 465
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1887 ATGATTAACAGAGAGACTAGAGAGACTTAA...AAGAAAGAGATCA 1933
466 AspAlaAlaGluValGluLeuSerLysPheAsnLeuSerArgIleAsp 482
1934 AATGACCAAAACAGTAAAGAACTTATGCTTCAATGAAGATTCCTCTGCA 1983
482 uArgGluArgHisMetGlnIleGluVal..... 491
1984 AGACAAAGAGATATGTACGCTGAAGTAAAGCTCTTATAGAGAAACA 2033
492 .....GluArgLysThrLeuAlaLeuGlyGluArgAspTyr 503
2034 ATAAACCTAGTTCAGAAAAAACAAGCTTGAGT.....AGGATTTG 2074
504 AspSerIleIleSerGlnLys.....ArgThrGlnValTyr 515
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2075 GAGGTATTTTGTCTCAAAAAGAAAGATGTATCTTAAAGACATATTAAC 2124
515 rSerLeuGlnGlnLysIleLysValLeuLeuArgGluLysAspIleIle 532
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2125 TCAATTAAGAAAAAGAACTTCAGTTAAATGCTTGAACAGCAAGTATTTAA 2174
532 snArgAsnAlaAsp...GluArgVal...LysLeuGlyLeuLysLysAsp 546
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2175 ATAAACCTCTTGAATAATGAGCAAGTTCAGAAAGTATTTGTTAAACCTG 2224
547 AlaLeuGluSerSerLysAspLysLysGlnIleVal.....AsnG 561
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2225 TTGTATGTTTCTTAAAGAAATGGGATCAGAAAGTTTCAGAAAGACAGTGA 2274
561 uHisLysAspLysIleLysLysValLeuArgGlyArgAsnProPheGlu 578
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2275 AGAGAAAGATGTTGTTAATGTCTTACAGGACGCTGTAATCTTGGCAA 2324
578 yAspMetLysLysGluIleAsnGlnAlaPheTyrProValAspLysGlu 594
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```


2325 AATAAATGAGAAAAATGCAACTGGCTTT 2356
595 TYRAsnGluLeuArgSerLysSerGlnAlaGlnGlnGluLeuLysPhe 611
2357 CAGCGTGATGAAAAAGTATTAGATTAGAAAAAGAGATTAAAGT 2400
611 eThrGlnSerLysValThrAspAlaArgGlnGlnLeuThrLysLeuArg 628
2401 CCTTCAGAGAGAGAGTGTAGTCTCAGTGTGACAGAACTTAAGTCTTTTTC 2450
628 rGAspMetAspAlaLysArgArgPheLeuAspSerLysLeuGlnSerIle 644
2451 GAGACTATGACGACGAAAGATCTCTTAAGGAAAGAGTTAGAA 2494
645 LeuGlnIleSerAlaAsnValAspMetPheProLysValLeuGlnAspAl 661
2495 GAAATTCAGCTCAGAAAAAGGCCCTGCAGCTGATCTTCTAGAA... 2539
661 aMetAsnLysArgAspGlnGlnLysArgLeuGlnAsnPheAlaAsnGlyM 678
2540 .ATGAGAAATGCTATGAAAAACAAGCCTTGAAT 2575
678 eTArgGlnMetLeuAlaProPheGlnHisLeuAlaArg Lys 691
2576 .CAGAAATCTTTAATTCAGTTGAAAGAGATCTCAACATGATGACAA 2623
692 AsnHisValLysProCysGlnLysArgAlaPheThrProAspGlnGlnAs 708
2624 AGTGAATCCATTAATGAAAAAGAAATGTTTATA AACGACA 2667
708 pGluPheValLysLysGlnArgMetGlnAsnSerSerThrAlaGlnArgS 725
2668 TGAAACCTTAAGCCACTACTAGACAAAAGAAATACGACATAGAGAG 2717
725 eTysAlaLeuAlaMetGlnSerSerAsnAlaGlnAlaLeuPheGlnGln 741
2718 CAGAGTTGATCTATTAAAGGATCTTACGAAATACACCTCTGTAAAA 2767
742 LeuAspLysLeuArgThrIleTyrAspAlaIleTyrValLysLeuValGln 758
2768 AATGATCCTCTGCTTCA GTAAAGAGCTGTGAAGA 2802
758 uThrIleProLeuAlaGlnLysAsnLeuAsnGlnHisLeuAlaAspGln 775
2803 AAAAAATA GAAATCTGAAAAAGAAATGCAAGAAAGG 2840
775 eGlnLysAlaGlnAlaPheAspLysLeuGlnLysValLeuAlaHisVal 791
2841 AGAGAAATAATAATAGATA... AATTAAGTTGCCGTAAGCAGAAAGAA 2887
792 GlnMetAspArgAspAlaValGlnAlaLeuLeuGlnProThrAspThrI 808
2888 GAACATGATTCACAGCAAAAGAG ACCCAGACGTGT 2922
808 eAspArgHisValHisGlnIleGlnGlnLeuValLysGlnValGlnAsp 825
2923 GAAG GAGGAACTTGATCTCTTCGATCAGAAAAAGGCCGT 2963
825 euGlnTyrAlaLeuAspSerSerGlyArgGlyValLysSerLeuGlnGlu 841
2964 TATGTGCTTCATGAGAGATCTCATTCAGAGACGAAAGCTATTAAAGAT 3013
842 IleGlnLeuGlnLysAsnPheLeuGlnArgThrArgAspThrLeuIleVal 858
3014 CTTTATTATGAA TATGAAAGCAGTCAGACAACTGAGAGT 3054
858 IglValAlaAspLysLeuArgAsp... GlnHisArgMetLeuAsnGlnAsp 874
3055 GGAAGAAAGAACGTCTAATAATTTTGAAGCATCGTAAT GAGAGCC 3098
874 et SerSerAlaGlnValArgThrPheHisAsnAla 884
3099 TTACAGACACATTAAGAAATTCGACTTTCGAGTGTGAACAAATAATCTCT 3148
885 ArgGlnGlnLysValLysAlaSerSerIleLeuGlnLysArgPheGlnLys 901
3149 GATTAATGAAGATCTCCTGCTGCT ATGAGACATTTACAGCTTAA 3192
901 rGlnGln GluLeuValLeuLeuAlaGlnGlnLysGlnGlnLeu 916
3193 TGCCAAATATTATAGAGTACAGATTTTAAAGATCCGAGAGACCAAGCA 3242
916 LeuAlaGlnLysLysLeuLeuGlnGlnLysSerLeuAspProLeuSerLysGln 932
3243 TGTATACCAAGAAATTAAGAGCTGAAAACTT CAG 3277
933 LysGlnSerLeuLeuGlnGlnLysTyrAsnAlaLeuLysGlnLysLeuAsp 949
3278 AAGAACAGAGATTAAGAACATGCGACTACTGTAAAGAACTGAAAG 3327
949 uGlnTyrHisGlnLeuAlaGlnLysArgGlnPheGln G 963
3328 ACTTCAGTACAACTTCAAAAGAAAGAAACAGCTTCAGAAACCATGC 3377
963 IglLysLeuAspAlaLeuGlnArg LeuAsn 972
3378 AAGAAATTAAGCTGTTAAAGAGATGCCCAACAAACCATTTGATGAT 3427
973 MetLysIleLysGlyTyr LeuAspSerLys... 982
3428 ATGGAATAGCTGATTAATGACGTTTGATGAAGAACTTAATCAAAAGTT 3477
983 LysAsnGlnLysLeuLysGlnLeuGlnGlnLysValLeu 997
3478 AACTAATTAACAAACAGATGAAAGATTTGAG 3511
997 uHisSerGlnLeuGlnSerCysMetAlaLysGlnGlnArgIleSerAla 1013
3512 CAGAAATTAATAATTCAAAACAGAAACAAAGAAACCTTACAGAA 3556
1014 GluLeuAsnLysSerLysGlnLeuGlnGlnGlnGlnLysGlnLeuLysArg 1030
3557 GAAATTAAGCTTCACTACAGCTTCAGTACACAAATTAAGAAAGAAAC 3606
1030 gAsnIleAspAsnLeu... LysTyrArgGlyThrLysAlaAspValG 1046
3607 CAATAATCAAGCAATGCTGTGAAACCAAAAGAAAGAACTGCGAGTTGCA 3656
1046 IglLysLeuThrArgAspIleGlnSerLeuGlnGlnLysArgLeuSerIle 1062
3657 AGCAAGCAGAAACGTACATTAATCTTCAAGCATCTTTAAA 3700
1063 GlySerLeuSerAlaIleGlnLysLeuLysArgHis 1075
3701 GGTGAGCTGAGAGCAGACGACGACAGATGAAAGCTTAAATACAGCT 3750
1076 SerGlnGlnLys GluArgLeuAsnSerG 1085
3751 GCGTGAATTAACATCAGAGAGCAGCAAAATCCACGACGACCTGAAACCT 3800
1085 IupPheAsnArgTrpGlnGlyThrLeuSerValTyrGlnSerAsnIleSer 1101
3801 CTGCGGACAGACGACGCTGAGCTGAGTGCATACACAGAGAGAGTACA 3850
1102 LysHisLysGlnGlnLysLeuSerGln 1111
3851 GCACACAGAGAAAGATGCCGTGCTCCAAAGCAGAAACAGCTACTGTAC 3900
1112 TyrLys AspIleGlnLysArgT 1119
3901 CTCTGAATTCAGAGCTCAAAAGTCCGAGTCAATGTTCTTAAACAAAC 3950
1119 yThrAsnGlnPheLeuGlnLeuLysThrThrGlnMetAlaAsnLysArg 1135
3951 AGAAAAATTAATCTATGCTGACGCTGAAACTGAGGGCGCTAAACAGAA 4000

```

1136 leuaparcTyyTyrhaleuap ..... 1144
      :::::
4001 AGGGAACATCTGGAAATGCTGATTGCACAGCTAAAAATCAATTACAGA 4050
      :::::
1145 ..... Lysaleueta 1149
      :::::
4051 TAGCCAAATTACTTACAGATTATGTATCTGAACTCAAAACATTTGCAGT 4100
      :::::
1149 raphHisSerMetLysMetGluGluIleasnLysIleIleLysGlu 1164
      :::::
4101 CTGAACATGATACACGTCTAGAAAGGCACACACAAGATGCTGCAGGAA 4147
      :::::
seq_name: /SIDS1/gcdata/geneseq/geneseqn/NA1998.DAT.ANV21209

```

```
seq_documentation_block:
ID   AAV21209 standard; DNA; 1664976 BP
vv
```

AC	AAV21209;
AD	10-NOV-1998 (first entry)
AE	Methanococcus jannaschii circular chromosome.
AF	Methanococcus jannaschii; methanogenic archaeon; circular chromosome;
AG	genome; autotrophic; extrachromosomal element; identification; ds.
AH	Methanococcus jannaschii.
AI	MO9807830-A2.
AJ	26-FEB-1998.
AK	22-AUG-1997; 97MO-US14900.
AL	22-AUG-1996; 96DS-0024428.
AM	(GENO-) INST GENOMIC RES.
AN	(UNIT) UNIV ILLINOIS FOUND.
AO	(UJJO) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
AP	Balt CJ, Smith HO, Venter JC, White OR, Woese CR;
AQ	WPI; 1998-169145/15.
AR	Complete genome sequence of methano-genic archaeon, Methanococcus
AS	jannaschii - useful in identification of M. jannaschii genome
AT	fragment
AW	Claim 13; Page 152-585; 614pp; English.

The present sequence represents the complete 1.66-megabase pair genome sequence of the *Methanococcus jannaschii* circular chromosome. The present invention describes M. jannaschii open reading frames from the genome sequence. The invention also describes a computer based system for identifying fragments of the M. jannaschii genome that are homologous to target nucleotide sequences, comprising: (a) data storage means comprising the nucleotide sequence of the 1664976, 58407 or 16550 bp sequence (see AA121209, AA121210 and AA121211), or a nucleotide sequence at least 99.9% identical to it; (b) search means for comparing a target sequence to the nucleotide sequence of the data storage means to identify a homologous sequence, and (c) retrieval means for obtaining the homologous sequence. The method, which is based on whole genome random sequencing of an autotrophic archaeon M. jannaschii, the genome of which consists of 3 physically distinct elements, a large circular chromosome (the 1664976 bp sequence given in AA121209), a large circular extra-chromosomal element (the 58407 bp sequence given in AA121210), and a small circular extra-chromosomal element (the 16550 bp sequence given in AA121211), can be used in the identification of M. jannaschii genome fragment.

Sequence, 1664976 BP; 568133 A; 266649 C; 256701 G; 573392 T; 101 other;

```

alignment_scores:      Length: 1366
                        Quality: 465.50
                        Ratio: 0.721
                        Gaps: 54
                        Percent Similarity: 47.640
                        Percent Identity: 18.985

alignment_block:
US-09-538-396-2 x AAV21209 ..

```

alignment_block:
US-09-538-396-2 x AAV21209

Align seg 1/1 to: AAV21209 from: 1 to: 1664976

7 Metenilleuclgyllylearg SerPheaspProaspbnlysys 21
198975 ATATATCTATAAAGAAATTAAGATGAATTAAGCTTAAGAAAGCATGTGTAATTC 199024
21 valIlethrPhePheLyPheProLeuthrLeuIleValGlyProAsnGlyA 38
199025 AAGAAATTAAGCTTGAAGAAAGCATGTGTTCGAAATTAATTTGGAGAAATGGAA 199074
38 laclLystrThrIleIleGluCysLeuLysLeuSerCysThrGlyGlu 54
199075 GTGGAAATCATCTATCTTGAAGCAGGTGTCCTTCCTGTTGGG... 199121
55 LeuProAsnSerIatSerGlyLysIleThrPheValHisAspProLysVal 71
199122GCAGCCAGTAATTTAATTAAGCACCAATTAAT 199153
71 lAlaGlyLutThrGluThrLysGlyLnIleLysLeuA 84
199154 A.....ACCAAGAAAGAAAAATCCGTTAATGTTGAATGG 199188
84 rghelystrhAlaAlaGlyLysAspValAlaCysIleargSerPheGln 100
199189 ATTTTGA...GTCATGGAAACACACACAAATTTATCAGAGATATGAT 199235
101 LeuthrGlnLysAlaSerLysMetGluPheValAlaIleGluSerValle 117
199236 TCTGGAAAGAGGGGAGCTAAAGCTC...TATTAAGAAATGGAAAGCCTTAAGC 199282
117 uGlnThrIleAsnProHisThrGlyLysValCysLeuSerTyrArgC 134
199283 TACACACATTAAGCA..... 199298
134 ysaAlaaspMetAspArgGluIleProAlaLeuMetGlyValSerLysAla 150
199299GTTAATTAAGCAGTAATGCAATCTTAAGCCGTGATAGAAAT 199340
151 ValLeuLysAlaValIlePheValHisGlnAspGlu..... 162
199341 ATGTTCTTAACCTCCATATATATTAACAAGGGGAGATACCTAAATTTT 199390
163 SerAsnTrpProLeuGlnAspProSerThrIleLysLys..... 175
199391 GAGTTTAAACCTCCGCAAAAAATTTGAAACAGTTGGCAAACTTTGGCA 199440
176LysPheaspAspIlePheSer.....AlaThrArg 185
199441 TACATGAGTTTGAAGAAATCTATCAAAAAAATGGGGAGATGTTAAGCA 199490
186 TyrThrLysAlaLeuGluValIle..... 193
199491 TATGAAAAAAGATTAAAGAAATTAAGAGAGAGAGCTGAATTAACAAGAAA 199540
194LysLysLeuHisLysAspGlnMetGlnIleLysThrPheArgL 209
199541 TTATGAAAAAGCAATTAATAAAATAAATGAGACCAATTGAAGAA..... 199588
209 euLysLeuGlnAsnLeuGlnThrValLysAspGlnAlaHisLysLeuArg 225
199584 .AAAAATTAATAAATTAATGCAATTAATCATGATTAACCTAAATTAATAAATAAAA 199631
226 GluAsnIleAlaGlnAspGlnGluLysSerAspAlaSerLysSerGlnMe 242
199632 AAGGAATTT...GAAGTATATGAAAAATTTATTAATGAAGAGGAAAAATAA 199678

242 tGluGlnLeuLysGluLys...IleCysGlyThrGluArgGluIleLeuG 258
::: ||| ||||| ||| :::::
199679 AAGTGTGTGTATGAAAAATTCATAAACAACCTTGAAGAAAAGAGAGAG 199728
258 LInetGluThrSerLeuAspGluLeuArgArgLeuGlnGlyGlnIleAsp 274
::: ||| :::::
199729 CTTTAGAGCTTAAAAATCAAGAGCTTAAAAATTTAGATAATGATTAAAT 199778
275 LLeuSAlaThrGluArgSerThrLeu...LeuThrGlnGlnHisG 289
::: ||| :::::
199779 ACCTGTGTGAGCAAGAGAACTCTAAATAGACATTAAGATGAATATGA 199828
289 uLysLeuAlaLeuSerGluGlnGlnGlnAspThrAspGluGluLeu 306
::: ||| :::::
199829 AAAATATAATCATTAAGTGTGATGAGATTAGGAAAAATTGAGAGCAGATTAA 199878
306 etGluTPGlnThrLysPheGluGlnArgIleAlaLeuLeuGluThrLys 322
||| :::::
199879 GAGATTAAGAGCTCATTAAGACATTATTAAAA... 199913
323 LLeuSerLysLeuValArgAspMetAspAspGluAlaSerTyrSerSerVa 339
199913
339 LLeuSerLysGlnAsnSerGluLeuThrHisGluIleGlyLysLeuGlnA 356
||| :::::
199914 TTAACAAGACAGCTTGAGATTAATAAAGAGACATTCGAAAAATTTGAAA. 199961
356 laGluAlaAspAlaHisLeuThrMetLysHisGluArgAspSerAspIle 372
::: ||| :::::
199962 ...GAAATTCATCAACAAGAACTAAAGTATAGAGATGATTT. 199997
373 LysAsnIleCysThrLysHisAsnLeuGlyProValProGlnHisProPh 389
||| :::::
199998 GATAAATTAGTACT... 200012
389 eThrAsnAspValAlaMetAsnLeuThrAsnArgIleLysAlaArgLeus 406
||| :::::
200013 ...CTATTAAATTAATAAATAAAGATGAAATTTG 200040
406 eTserLeuGln...AsnAspLeuAspLysLysSerAsnGlu 420
::: ||| :::::
200041 AAAGAGTGAAGACTTAAAGATTGCTTGAAGAACTTAAATAATTAAT 200090
421 AspGlnLeuAspValLeuTrpLysHisTyrLeuLysIleAsnAlaArgTy 437
::: ||| :::::
200091 GAAGAGATTGA... 200102
437 rSerGluValAspGlyGlnIleGlnSerLysIleGluSerMetSerGlyI 454
||| :::::
200103 ...AAAATGAAAAAATATAAAGAA 200124
454 lLeuAlaArgArgArgLys...AspLysGluLysGluArgAspAla 467
||| :::::
200125 TATGTGAAGAGCTGCAAGAAATCTATGAGAAATATTGAAATTAAGAA 200174
468 AlaGluValGluLeuSerLysPheAsnLeuSerArgIleAspGluArgG 484
||| :::::
200175 AAGCTGTAGAAATACATAAATACTACTTA... 200204
484 uArgHisMetGlnIleGluValGluArgLysThrLeuAlaLeuGlyGluA 501
::: ||| :::::
200205 GAGATATATACATTGCTTCAGAGAAAAAATCCATGAAAAAATAATTA 200253
501 rGAspTyrAspSerIleIleSerGln...LysArgThr 512
::: ||| :::::
200254 ACGATTATGAACAACAGAAATTAATAAATCTTTGAAGAAACAATAATATT 200303
513 GluValTyrSerLeuGlnGlnLysIleLysValLeuLeuArgGluLysAs 529
::: ||| :::::
200304 GACATTGAAGATTTGAAATTCATTAAAGAGATAGAGAAAAAAGAA 200353

529 pIleIleAsnArgAsnAlaAspGluArgValLysLeuGlyLeuLysLysA 546
::: ||| :::::
200354 AGTCTTGAATAATCTACAAAAAGAAAGATGAGACTTAACCAAAACCTTG 200403
546 sPAlaLeuGlnSerSerLysAspLysLeuAsnGluIleValAlaGlnHis 562
::: ||| :::::
200404 CGCAAAATTAACAGTGAATTTAAAGGCTGAAAAAATTTTGAAGAACTT 200453
563 LysAspLysIleLysLysValLeuArgGlyLysArgProPheGluLysAs 579
||| :::::
200454 AAAAGAGTTGAGGAAAAATGCCACTGTATGAAACCCCAATAGTGAATA 200503
579 pMetLysLysGlu...IleAsnGlnAlaPheTrpProValAspLysGluT 595
||| :::::
200504 TAAAAAGTGAATTAATTAATCAACATAAACCAGGTAAACATAAT 200553
595 TyrAsnGluLeuArgSerLysSerGlnGluAlaGluGlnGluLeuLysPhe 611
||| :::::
200554 ATACGATTTGACGAA... 200570
612 ThrGlnSerLysValThrAspAlaArgGlnGlnIleuThrLysLeuArgAr 628
::: ||| :::::
200571 ATAAATTAATAAATAATGAGAGATTGAAAAAGATTATAGGAAATTAAGAA 200620
628 GAspMetAspAlaLysArgArgPheLeuAspSerLysLeuGlnSerIleL 645
::: ||| :::::
200621 AGAAATTTAT...AAAGAGAAATCTTAAGACACTAAGACCTTAATTC 200667
645 eugInIleSerAlaAsnValAspMetPheProLysValLeuGlnAspAla 661
||| :::::
200668 TTGAATAACAAGTCAGATGTAAGAATTAAGAAATTAATAAATTAAT 200717
662 MetAsnLysArgAspGlu...GlnLysArgLeuGlnAsnPheAla...As 676
::: ||| :::::
200718 AAAGAGCAGTTAGAGAAATCAATTAATAAATATCCCACTGTGTAATTA 200767
676 nGlyMetArgLysMetLeuAlaPhePheGlnHisLeuAlaArgLysAsnH 693
||| :::::
200768 CGGA...AACCCAGTGATGAGATA... 200789
693 lSValCysProCysGluArgAlaPheThrProAspGluGluAspGlu 709
200789
710 PheValLysLysGlnArgMetGlnAsnSerSerThrAlaGluArgSerTy 726
200789
726 sAlaLeuAlaMetGluSerSerAsnAlaGluAlaLeuPheGlnGlnLeuA 743
||| :::::
200790 ...TTAGAAACACTTAAAGCCAGCTGA 200814
743 sPLeuLeuArgThrIleTyrAspAlaTyrValLysLeuValGluGluThr 759
||| :::::
200815 ACAAAATTAATAAATCTTATATACCAATACTATGCTGATGATTA 200864
760 lLeuProLeuAlaGluLysAsnLeuAsnGlnHisLeuAlaAspGluSerG 776
::: ||| :::::
200865 AATAGCTGATGATGAGGAAAGT... 200885
776 nLysAlaGlnAlaPheAspAspLeuLeuGlyValLeuAlaHisValGlnM 793
200885
793 eTAspArgAspAlaValGluAlaLeuLeuGlnProThrAspThrIleAsp 809
200885
810 ArgHisValHisGluIleGlnGlnLeuValLysGluValGluAspLeuG 826
||| :::::
200886 ...ATTGAATATAGAAATTAAGGAAATTAAGAAATATGCT 200920
826 uTyrAlaLeuAspSerSerSerGlyArgGlyValLysSerLeuGlnIleG 843

XX (PRIV-) PRIVATES INST IMMUNOLOGIE & MOLEKULARGEN.
XX
XX Renz M, Seelig HP;
XX
XX WPI: 1996-403153/41.
DR P-PDB: AAM02258.
XX
XX DNA encoding nucleolar-endosomal auto-antigen - useful for exact
XX diagnosis of rheumatic disease, in gene therapy and for removal of
XX specific auto-antibodies
XX
XX Claim 2, Fig 2, 15pp. German.
XX
XX Transformed cells can be cultured to produce the antigen p162, for use
XX in exact (differential) diagnosis of rheumatic disease, i.e. they
XX can detect, in immunoassays, Western blots, etc., rheumatism-
XX specific auto-antibodies. The antigen can be used therapeutically,
XX in the removal of auto-antibodies from the circulation, or when
XX coupled to a cytotoxin, the elimination of auto-antibody-
XX producing lymphocytes.
XX
SQ Sequence 4779 BP: 1916 A: 801 C: 1007 G: 1055 T: 0 other:

alignment_scores:
Quality: 455.00 Length: 1364
Ratio: 0.654 Gaps: 60
Percent Similarity: 51.026 Percent Identity: 21.628

alignment_block:
US-09-538-396-2 x AAT58751 ..

Align seg 1/1 to: AAT58751 from: 1 to: 4779

135 AlaasMetaspargluileproalaleumeclyValserlysAla 151
122 TCTGATTAGAT.....TCATCGCAAC 144
151 lleugluAnValilepheValHleuglnaspgIuserAsnTrpProleug 168
145 TCCTATAACACAGTGCAGCTCATATAGAACCTCTCAGAGGTTCA 194
168 lnaspproserThrLeuLysLysLysPheaspIlepheserAlaThr 184
195 TATGTCCTCCAGTATGAAA.....TCTCTTGATCTGCTGAT 232
185 ArgTyrThLysAlaLeuGluValileLysLysLeuHisLysaspIle 201
233 GAACCTTTCAACATTATGAACTGTTCAATGATCTGTAATGACTCAGG 282
201 tGlnGluileLysThrPheArgLeuLysLeuGluAsnLeuGlnThrVal 218
283 TCTGAGAGGAGAGCT.....AATCTTGCCTTGAGC 314
218 ysaspIleAlaHisLysLeuArgLysAsnIleAlaGlnaspgIleLys 234
315 GAGATGATGTAACAGCTCAGACAGAGGCTCAAGACCTACAGCTCA 364
235 SerAspAlaSerLysSerGlnMetGluGlnLeuLysGluLysIleCys 251
365 CTTTAAGGAAGAAAAGTACTCGAAGAAATTAAAGAG..... 403
251 yThrGluArgGluileLeuGlnMetGluThrSerLeuAspGluLeuArg 268
404GAATTAGAAA 413
268 rglenglnGlnGlnleAspIleLysAlaThrGluArgSerThrLeu 284
414 AATATCAAGAGG...CTGACAGACAGAGAGCCAAACCTGATGGGTG 460
285 ThnglnGlnIsgLysLeuAlaAlaLeuSerGlnGluAsnGluAspTh 301

461 ACTGATTCATCAGACAGAACTACAGCTCTTGGACACCAATTAGAGAGC 510
301 raepgluGlu.....LeuMetGluTrpGlnThrLysPhegluGluA 315
511 CCAAACAGAAAATTTAATATTAAGCAAAATGAAAGACTTATTTGACAA 560
315 rglleAlaLeuLeuGlnLutThrLysIleSerLysLeuValArgaspMet 331
561 AACGAGCCCACTTCTGCTACTGAAATTCAGATATTAAGTCAAAGTGTG 610
332 AspGluAlaSerLysSerValleuSerLysGlnAsnSerGluLeuTh 348
611 GAGAAAGAGCTCTTCGAGAGCTGCTGAACAAAGAGTACAGCTGAC 660
348 rHsGluileGlyLysLeuGlnAlaGluAlaAspAlaHisLeuThMet 365
661 AGAAGAAATTAAACAA.....GAGGCACTGTAATTCACAGACTGA 701
365 yHsGluArgAspSerAspIleLysAsnIleCysThrLysHisAsnLeu 381
702 AGACGGAA..... 709
382 GlyProValProGluHisProPheThrAsnAspValAlaMetAsnLeu 398
710CTGCTTCAGAGAGCTGCTGATAGAAAGATGTCGCGTG...CTAAA 750
398 rAsnArgIleLysAlaArgLeuSerSerLeuGlnAsnAspLeuAsp 415
751 GAAAGAACTGCTCCAACTTCAACACTAATGATTAACATGACCTTGAAAC 800
415 yAsLysSerAsnGluAspGlnLeuAspValLeuTrpLysHisLys 431
801 GTGAGCGAGATCTGAAAACCTCAAGAT.....GAATGCAAA 838
432 LysIleAsnAlaArgTyrSerGlnValAspGlyGlnIleGlnSerLys 448
839 AATTCGACGACATATGCTAGCTCAGAGCCCAATAGCCAG..... 883
448 egluserMetSerGlyLeuArgArgArgLysAspLysGluLysGluA 465
884CTAAGGAGTGAACCTGCCAAAGGCCCCACAG 914
465 rgaAspAlaAlaGluVal...GluLeuSerLysPheAsnLeuSerArg 480
915 AAGTTCGTATATGTACAGAGAACTCAAAACCTGAAA...AATTCAGTT 961
481 AspGluArgGluArgHisMetGlnIleGluValIleArgLysThrLeu 497
962 AATGAAATTAACACAAAAAATCAGACCTTGACAGAA.....AATTCGT 1005
497 AluGluGluArgAspTyrAspSerIleIleSerGlnLysArgThrGlu 514
1006 GAAAAAAGAACAGACTATCTAAG...TTAGAGGAGAAACATTAATGAG 1052
514 AlTyrSerLeuGlnGlnLysIleLysValleuLeuArgGluLysAsp 530
1053 AATCTGAGCTAAAGAAATTTCAAGCACCCTTCATCAAAAAGACCTA 1102
531 IleAsnArgAsnAlaAspGluArg..... 538
1103 GATTGTCAACAGCTCTCAGTCAAGATTCTCGCATCTGAAACCTCAGTCA 1152
539ValLysLeuGlyLeuLysLysAspAlaLeuGluSerSer 552
1153 TAGAATACATGTAAGTAACTAGTGAAGAAAGAGAACTACTCAAAAGCTCA 1202
552 ysaspLysLeuAsnGluileValAsnGlnHis.....LysaspLys 565
1203 AAGAAAGATATCTGAGGTAGAGACCAAGTACCAAGATCTAAAGGGGAG 1252
566 IleLysLysValleuArgGluArgAsnProPheGluLys.....As 579
1253 TTTAAGCAGCTACACACAGAGAGAAAGAAAGAGCAGCATGGGTACA 1302


```

3025 TGTATTACAGACAGACAGCTGAGATATAACTACAGACAGCTTAACAC 3074
1078 .....GluLysGluArgLeuAsnSerGluPhe 1086
3075 AGGACAGCCAGAACTTGACAGAGAAAGAAA..... 3109
1087 AasnArgTrpGlnGlyThrLeuSerValTyrGlnSerAsnLeuSerLysH 1103
3110 .....ATATCAGATTATCAAAACAACTATGAAAAAAG 3141
1103 slysglnGlnLeuLysSerGlnTyrLysAspIleGluLysArgTyrT 1120
3142 TCAGGAA.....ACTTCAAAACAGCTTCATCGATTCT 3176
1120 hAsnGlnPheLeuGlnLeuLysThrThrGluMetAlaAsnLysAspLeu 1136
3177 ATGGGAGGAAATCTGAACTTACGACACAGCAAGATCTTAAGTCTGTA 3226
1137 AspArgTyrTyrThrAlaLeuAspLysAlaLeuMetArgPheHisSerMe 1153
3227 GAAGAGAAAGCTTCTCTACACAGAGAGACTTGATTCA...AACAGAAA 3273
1153 tLysMetGluGluIleAsnLysIleLeuLysGlnLeuTyrPglnglThrT 1170
3274 TCAAATGGAATCAAAATAATATGATTCAAGACTG.....AAGACTG 3317
1170 YrArgGlnAspIleAspTyrIleSerIleAsnSerAspSerGluGly 1186
3318 CCAAGGCT.....ACATGGAGCAGAGATTCA..... 3343
1187 AlaGlyThrArgSerTyrSerTyrArgValValMetGlnThrGlyAspAl 1203
3344 .....GCAAGAA 3351
1203 aGluLeuGluMetArgGlyArgCysSerAlaGlyGlnLysValLeuAlaS 1220
3352 AGAACAGCATGTCAGAGCAGATGTAAGCATCAAGACATTCAGAAAG 3401
1220 eLeuIleIle.....ArgLeuAlaLeuAla 1228
3402 AAAGCATCATGAAAGAAAGAAAGAACTGTAAATGAGAAAGCTAAATTGGCA 3451
1229 GluThrPheCysLeuAsnCys.....GlyIleLeuAlaLeuAs 1241
3452 GAGATGAGAAATTAATATGTAGACAAAGAAAGAAATCACTAACTAAA 3501
1241 pGlu.....ProThrThrAsnL 1247
3502 CGAAGAACTCAAGTCCCAAACTAGAAAGCATTAAGAGATTAACAATTC 3551
1247 euAspGlyPro..... 1250
3552 TTAAGATGCTAAACAGCTTCTAATTCAGAGAAATTAGAACTCAAGGA 3601
1251 AasnAlaGlnSerLeuAlaAlaLeuLeuArgIleMetGlnAlaArgGly 1267
3602 AAAGCGAGACTCCCTGAAGCAGCT.....GTTGAACAGAGAGAAAGAG 3642
1267 sGlyGlnGlnAsnPheGlnLeuIleValIleThrHisAspGlnArgPhea 1284
3643 AAATCAGCAGATTAATAAGACCAAGGTGAAGAAAGAAAGAG..... 3685
1284 lAsnIleuIleGlyGlnArgGlnLeuAlaGluLysTyrTyrArgValSer 1300
3686 .....GAGCTGAAGAAAGAAATTT.....ATTGAG 3709
1301 LysAspGlnAsnGlnHisSerIleIleGluSerGlnGluIle 1314
3710 AAAGAAGCTAAGTTGCATTCGGAATAAAAGAAAGAAAGAACTA 3751
seq_name: /SIDSL/gcgdata/geneseq/geneseq/NAL1999.DAT:AXX26819

```

```

seq_documentation_block:
ID AAX26819 standard; DNA; 9626 BP.
AC AAX26819;
DE 22-JUN-1999 (first entry)
XX
XX Nucleotide sequence of centromere-associated protein-E (CENP-E).
KM CENP-E; centromere-associated protein-E; ATPase activity;
KM plus end-directed microtubule motor activity; chromosome congression;
KM microtubule binding activity; chromosome movement; mitosis;
KM cell proliferation; tumor; metastasis; vascular malformation;
KM inflammatory disease; immune disease; angiogenesis; hypertension;
KM resclerosis; fungal infection; selective herbicide; fungicide;
KM insecticide; plant growth regulator; activator; cancer cell marker; ss.
XX
XX Xenopus sp.
OS
XX WO9913061-A1.
XX
XX 18-MAR-1999.
XX
XX 10-SEP-1998; 98WO-US19231.
XX
XX 11-SEP-1997; 97US-0058645.
XX
XX (REGC ) UNIV CALIFORNIA.
XX
XX Cleveland DM, Goldstein LSB, Sakowicz R, Wood KM;
XX
XX WPI. 1999-229233/19.
XX
XX P-PSDB; AAY01632.
XX
XX Centromere-associated protein-E and related nucleic acid
XX
XX Claim 13; Page 67-73; 77pp; English.
XX
XX The present sequence encodes CENP-E (centromere-associated protein-E)
XX of Xenopus. The protein has at least one of plus end-directed microtubule
XX motor activity, ATPase (adenosine triphosphatase) activity and
XX microtubule binding activity. CENP-E is the motor that powers chromosome
XX movement toward microtubule plus ends and is essential for congression
XX of chromosomes during mitosis. Modulators of CENP-E can thus control
XX cell proliferation. Agents that modulate CENP-E activity are lead
XX therapeutic, bioagricultural and diagnostic agents, e.g. for treatment
XX of unwanted cell proliferation (typical of many examples are tumors and
XX metastases; vascular malfunction; inflammatory and immune diseases;
XX angiogenesis; hypertension; restenosis; and fungal infections), also as
XX plant protection agents (selective herbicides, fungicides and
XX insecticides) and plant growth regulators or activators for improving
XX yields. CENP-E is also a diagnostic marker for dividing cells, including
XX cancer cells.
XX
XX Sequence 9626 BP; 3540 A; 1692 C; 2104 G; 2290 T; 0 other;
XX
alignment_scores:
Quality: 449.50 Length: 1506
Ratio: 0.634 Gaps: 62
Percent Similarity: 47.078 Percent Identity: 20.053
alignment_block:
US-09-538-396-2 x AAX26819 ..
Align seq 1/1 to: AAX26819 from: 1 to: 9626
100 GlnLeuThrGlnLysAlaSerLysMetGluPheLysAlaIleGlnSerVal 116
2951 GAATTGAGTAAAGAAAGTTCTGAGTTCTCA...AACACAGCTGAGAAAC 2997
116 lLeuGlnThrIleAsnProHisThrGlyGluLysValCysLeuSer...T 132

```


2998 ATTGAGAAAAAAATGCTTGGAGATGAAGTACCTTGCCTATCAGAAAT 3047
132 yTArGcysAlaAspMetAspArgGluIleProAlaLeuMetGly...Val 147
13048 ACAAAATTT...TTGCCAAATGAAGTGAATGCTTGAATAAATCAGATC 3091
148 SerIysAlaValLeuGlu... 153
3092 AGCAAGGCTTCTGAAGAGATTAATGTTAAGCAAGAGAGAGAACTTC 3141
154AsnValIlePheValHisGlnAspGlu 163
3142 TGCATCTATTATTAAGCAAAACAGATTAATCATCAGAGACAACTCAGC 3191
163 eTAsnTTProLeuGlnAspProSerThr... 172
3192 AGATTTTAAACAGTACAGCAAGTGAACACACAGTCAAAAGTCAG 3241
173LeuIysLysLysPheAspArgIlePhe 181
3242 CAGACGTAGAGCAATTAATCTGGAGATGAAGAAATGCATGATGATCTTT 3291
181 eSerAlaThrArgTyr...ThrIysAlaLeuGluValIle 194
3292 TGA...AAGTATATCAGAAACAAAGTGAAGCTGAAGACCTTTAA 3335
194 ySLysLeuHis...LysAspGlnMetGln...GluIleLysThr 206
3336 GAGAAATGGAGAACCTTAAGAGCACTATGAGTGTGGAAAGTAAAGTT 3385
207 PheArgLeuLysLeuGlnAsnLeuGlnThrValLysAspGlnAla... 221
3386 GCTGACACAAACATGAACCTTAAGAAACTTAAGGGTAAGAGCAACT 3435
221 221
3436 GCTTCATGAGAAAAAATCTTTTCAAGCAATGACACTATATTCGA 3485
221 221
3486 TTACACCTTTTCAAGACTCGCTCCCTCCCAAAATTAAGTTGAAGGAAAC 3535
222HisLysLeu... 224
3536 TCTCAAGACCCCATAGAATCAATGACTACACATTAATTAAGCCCTTGC 3585
225ArgGlnAsnIle... 228
3586 TACAGAAAGAAACACATTATGCTGTCTAGAGACTGAAGAAACAGTC 3635
229Ala 229
3636 TCAAGAGCAAGTTATGATTGAAACACTCAACTCAAAAGCTTCAACGA 3685
230 GlnAspGlnGluLysSerAspAlaSerLysSer... 240
3686 CAAGTATTGAAGGTGATCTCCAGAAAGCCCAAGCAAGACTTGAAGA 3735
241GlnMetGlnGluLeuLysGluLysIleC 250
3736 AGGAGAGCTTAACTGCTTTTGGAGATGAACTACTGAAGGACACCTTA 3785
250 ySgIlyThrGluArgGluIle...LeuGlnMetGluThr...Ser 262
3786 CTGACTACAGCTGCTCTATAGAAATTTGCAAGCTGAGAACTGTGAAGTT 3835
263 LeuAspGluLeuArgArgLeuGlnGlnIleAspIleLysAlaThrGlu 279
3836 ACAGAAAAACTCCAAACACTTCACAGAGAGATGAATAATTAATCTATGA 3885
279 uArgSerThrLeuLeuThrGln...GlnHisGlu 290
3886 AAGGATGAGCTTCAAGCAACCTTGAAGACCTGAAGACAGACATGATA 3935

290 ySLeuAlaIle...LeuSerGlnGluAsnGlu...AspThrAsp 302
3936 GCCTAAACAGACCTTATGTAACCACTTGAAGCACTGAATGAACACAA 3985
303 GluGluLeuMetGluIleArgIleThrPheGluGluArgIleAlaLeu 319
3986 GATGAATTAAGGCGCTGCCAGGAAGAGCTTAAGACACAGAAAGCACTGCT 4035
319 uGluThrLysIleSerLysLeuValArgAspMetAspAspGluAsp 336
4036 TGAATGCTTGAACACAGCTT...TTGATTTGCTGTCTAGAA 4076
336 yIleSer...ValLeuSerLysGlnAsnSerGluLeuThr 348
4077 TTTTCATCCAAACAGATGATGATGCTTAACCAAGAAAGTGTCTATTG 4126
349 HisGluIleGlyLysLeuGlnAlaGlu... 357
4127 GGTGAATTAATTTGTTTCAAAAGTGAATGCTGCTGCTGAAGATGA 4176
358AlaAspAlaHisLeuThrMet 365
4177 GCTCAAAACATCTTGAAGCATTAATTTCAAGTGAAGTGAAGTGAAGTGA 4226
365 yHisGluArgAspSerAspIleLysAsnIle... 375
4227 CTCAATGAAATCTGTGGAGGAGAAACCTTGAATCACAACAAACCTC 4276
376CysThrLysHisAsnLeuGlyProValProGluHisProPhe 390
4277 AATGGCTTGAAGAGATATTTGGCAATCTGAAGAAAGCAAGTGT 4326
390 rAsnAspValAlaMetAsnLeu...ThrAsnArgIleLysAla 404
4327 GAATGCCATGTTGGAGATCTAAAGAGAACACAACTAAGCTCAAGAAC 4376
404 rG...LeuSerSerLeuGluAsnAsp... 411
4377 AAGCAGAGAAATTTCTAAGAAATCAATTCAGTTTGAAGAGCTG 4426
412LeuLeuAsp... 415
4427 TTCAGTGGTTCACAGAACCTGATGATGATGATGATGATGATGATGAT 4476
415 sLysLysSerAsnGluAspGlnLeuAspValLeuTrpLysHisThrLeu 432
4477 GCTAAAGCAGCAGAAAGAGCTGGAATTAAGATGATGATTTTGG 4526
432 ySile...AsnAlaArgTyrSerGluValAspGlyGlnIleGlnSer 446
4527 AACTGTACAAACCTGCAACACCAATTAATTAAGGAAATTTGAAACT 4576
447 LysIleGluSer...MetSerGlyIleLeuArgArgLysAs 460
4577 CCATTGCAAGCTGACCAATGAGAGAGACAGCATTTATCGCGCTTCAAGA 4626
460 pLysGluLysGluArgAspAlaValGluLeuSerLysPheAsn 477
4627 AATGAGATTAAGTCTTGGGAAAACTTGAAGCAATCAGATTTATC 4676
477 eUerArgIleAspGluArgGluArgHisMetGlnIleGluValGluArg 493
4677 TGAAGAGATTCAGAAAGAAAGCTGAAGCTGTAACAACTTGA... 4723
494 LysThrLeuAlaLeuGlyGluArgAspTyrAspSerIleIleSerGln 510
4724AsnCTTCAGAA 4734
510 sArgThrGluValTyr...SerLeuGlnGluLysIle 522
4735 AGAGATGAGAGCTCGCTTCTATTAAGATGACTGCAACAGAGCTAG 4784

```

522 yValleuLeuArgLysAspIleIleAsnArgAsnAlaAspGluArg 538
:: ||||| ||||| ||||| ||||| ||||| |||||
4765 AAGCTTGCTGAGTGAAGAAATATTATTCTAAAGAGATATTGACACAC 4834
539 Valys..... 540
4835 CTAAAGCATCATTCAGATCTCAGCTCAGTGCAGAAACACAGCAGA 4884
541 LeuGluLeuLysLysAsp.....AlaL 548
||| ||| ||| |||
4885 GCTACAGTTGGCTAAGAAATCTTGCATCGCTCTCTGCAATTTGTCGA 4934
548 euGluSerSerLysAspLysLeuAsnGluIleValAsnGluHisLysLys 564
:: ||||| ||||| ||||| ||||| ||||| |||||
4935 TAACCTCAAGAAAGAAACCTCTGCGAGATTGTGTCATCTCTGGAAGA 4984
565 LysIle.....LysLysValLeuArgGlyArgAsnProheG 577
||| ||| ||| ||| ||| ||| ||| |||
4985 AAGATATTGTTATTAACTGAGAAATTCATCAAAAACTAATGAAACAGA 5034
577 uLysAspMet.....LysLysGluIleAsnGluAlaPheTTPProVal 592
||| ||| ||| ||| ||| ||| ||| |||
5035 AAAATTACTACATGAAGAAAGAACTTGAACAGCTCAGTTGAGCTAA 5084
592 sPLysGluTyrAsnGluLeuArgSerLysSerGluGluAlaGluGlu 608
||| ||| ||| ||| ||| ||| ||| |||
5085 AGTGGAGGTGAGACATCTGATGAGAGATGATCATCGAATCGAAGTCTCA 5134
609 LeuLysPheThrGlnSerLysValThrAspAlaArgGluGluLeuThr 625
||| ||| ||| ||| ||| ||| ||| |||
5135 CTGAGTCTCTGACATGAGAAACATGATGATGAAACAACTTCTTGC 5184
625 sLeuArgAspMetLysAlaLysArgArgPheLeuAspSerLysLeuG 642
||| ||| ||| ||| ||| ||| ||| |||
5185 TCTTAACAGCAGATGCAAGTATTACTCAAGAAAGAAAGAGCTGCAAC 5234
642 InSerIleLeuGlnIleSerAlaAsnValAspMetPheProLysVal 658
||| ||| ||| ||| ||| ||| ||| |||
5235 AAACCCATGAACTTAACAGCTGAGTGCACCTCTAAAGAGAAATATT 5284
659 GlnAspAlaMetAsnLysArgAspGlu..... 667
||| ||| ||| ||| ||| ||| ||| |||
5285 GAATTGGGTTGAATTTTAAATGAGCGCAGCAAAAGCCATAAGA 5334
667 ..... 667
5335 GCATGTCTGTAATGAGATTAAGAACTTGAGCAAGCAGACAGAC 5384
668 ..GlnLysArgLeuGluAsnPheAlaAsnGlyMetLysGluMetLeuAla 683
||| ||| ||| ||| ||| ||| ||| |||
5385 TTCAATGTGAGATAGAAAGCTTATGCAAAAGCTTAAAGATTAAGAGTCA 5434
684 ProheGluHisLeuAlaArgLysAsnHisValCysProCysCysGluArg 700
||| ||| ||| ||| ||| ||| ||| |||
5435 GCGCTGGAACTTTAAAGATCTGAA.....CAAA 5466
700 gAlaPheThrProAspGluLysAspGluPheVal.....LysLysGlnAla 715
||| ||| ||| ||| ||| ||| ||| |||
5467 ACTAATTATTCGAACCAAGAAATGGAATGTTATGCGAAATGAGAG 5516
715 tMetGlnAsnSerSer.....ThrAlaGluArg..... 724
||| ||| ||| ||| ||| ||| ||| |||
5517 AATTGAAAAAATAGCCAGAGACTGTAAATGCTGAGAGGAGCAGCTGCAA 5566
725 .....SerLysAlaLeuAlaMetGluSerSerAsn..... 734
||| ||| ||| ||| ||| ||| ||| |||
5567 GACGACCTAAGGAGAGTCTGAAATGCTCATTTGAAATCAAGATGATCT 5616
735 .....AlaGluAlaLeuPheGlnGluLeuAspLysLeuArgThr 748
||| ||| ||| ||| ||| ||| ||| |||
5617 AAGAAAGCTCAAGAAAGCATGTGACAGCAGCAAGAAATTAAGTTACAGAAC 5666
748 lArgThrAspAlaTyrValLysLeuValGluGluThrIleProLeuAlaGlu 764

```

```

5667 TG...ACGCCAGATTCTGCTCTCAGAGAAAGAAATCTCTCTTGCAA 5713
765 LysAsnLeuAsnGlnHisLeuAlaAspGluSerGlnLysAlaGlnAlaPhe 781
||| ||| ||| ||| ||| ||| ||| |||
5714 AATCAGATCTTTATATATGTTGCACTGTGAAGAAACTTAAGCGAAG 5763
781 eAspAspLeuLeuGlyValLeuAlaHisVal..... 791
||| ||| ||| ||| ||| ||| ||| |||
5764 AGATGACCTGAAACCACTGTAAGCAACACCTGTTCTCAGAAATGAAACTC 5813
792 .....GlnMetAspArg 795
||| ||| ||| ||| ||| ||| ||| |||
5814 TTAGTCTGCTTTAAAGAAAGAAATTTGCATTTGCAATGAAACAGAGAG 5863
796 AspAlaValGluAlaLeuLeuGlnProThrAspThrIleAspArgHisVal 812
||| ||| ||| ||| ||| ||| ||| |||
5864 GACAAAGCTGATGCTGCCAGAAACAAATATATATCAGAGAGATATC 5913
812 HisGluIleGlnGluLeuValLysGluValGluAspLeuGluTyrAlaL 829
||| ||| ||| ||| ||| ||| ||| |||
5914 AATATATAGAAAGACAGTACTTCAACAGCCACCAATTTAAAGAAATC 5963
829 euAspSerSerGlyArgGlyValLysSerLeuGluGluIleGlnLeuGlu 845
||| ||| ||| ||| ||| ||| ||| |||
5964 TTATGAAAGAGAGACTTTATCCAGTATGAGAGCACTGCTTGAAC 6013
846 LeuAsnPheLeuGlnArgThr.....ArgAspThrLeuIleValG 859
||| ||| ||| ||| ||| ||| ||| |||
6014 ACAGAACACCTTGAGGAACATGGAAGACAAAGACTTGCAATTTGGCTAA 6063
859 uValAspAspLeuArgAspGlnHisArg.....MetLeuAsnG 872
||| ||| ||| ||| ||| ||| ||| |||
6064 AATGAGCAGGAGAGATGAGACCTGCCAATAAAGTAAGTACCTTAAC 6113
872 LysAspMetSerSerAlaGlnValArg..... 880
||| ||| ||| ||| ||| ||| ||| |||
6114 AAAGATGCTCTTCTAGAAAGAACATCAATCAATTAATTTACTACTCA 6163
881 ..... 881
6164 AAGAAAGTGAGGTGAAAGAAAGACCTTCTACTCAGAGACCTCAAA 6213
888 sValLysAlaSerSerIleLeuGluArgPheGlnLysSerGluGluGlu 905
||| ||| ||| ||| ||| ||| ||| |||
6214 GCAGCAGCTCTTCTCCAGATGGAAGAGCTCAGAGAGCTTTAAAGACCA 6263
905 euValLeuLeuAlaGluGluLysGluGluLeuIleValGlu..... 918
||| ||| ||| ||| ||| ||| ||| |||
6264 AAGATTGCACTTGGAAGAGGCCGAGAGAGAGATTAAGTGAAGCTAAT 6313
919 .....LysLysLeu.....LeuGluGluSerLeu... 926
||| ||| ||| ||| ||| ||| ||| |||
6314 GAATTAAGATCTCACTGCTAGATCTCTCTCTAGAGAGAGATTTCT 6363
927 .....AspProLeu 930
||| ||| ||| ||| ||| ||| ||| |||
6364 TCAGAAATGCTACATTTTGAATGAACCTGTAAGCGAAAGGAAACCTCT 6413
930 euLysGluLysGluSerLeuLeuGlnGluTyrAsnAlaLeuLysGlnLys 946
||| ||| ||| ||| ||| ||| ||| |||
6414 GCCATTGGAAGCAGCAAGTGGTTTCAGAAATGGAAGCAGCTATCTGACA 6463
947 LeuAspGluGluTyrHisGlnLeuAlaGluArgLysArgGluPheGln 963
||| ||| ||| ||| ||| ||| ||| |||
6464 TTAAAGATGAGACCATTCATTTGCGCAATCAAAAGCGAAAGATGA 6513
963 nGluLeuAspAlaLeuGlyArgLeuAsnMetLysLys..... 976
||| ||| ||| ||| ||| ||| ||| |||
6514 AGCTTAATAATAATAGCCAGTCTCGCTGAAGAAATTAAGATCTGACA 6563
977 .....GlyTyrLeuAspSerLysLysAsnGlnLysLysLeuGlu 989
||| ||| ||| ||| ||| ||| ||| |||

```

```

6564 AACAGATGATGATTCAGAGATTCAAG.....GAATCTTCAGAGAA 6607
990 LeuGlnGlyArgHisVal.....LeuGlnHisSerGlnLeuGlnSe 1003
6608 ...CAGCTCCATCTAAGTGAAGATTATGATATGAAAGCTGAAC 6654
1003 rGysMetAlaLysGln.....ArgHisSerAlaG 1014
6655 TCAATGCTCAGCAACAGAAAGACATCAACAACTTCAGAGAA 6704
1014 LuLeuAsnLysSerLysGlnLeuGlnGlnGlnLysArg 1030
6705 AAGTTAAGGAAGTGATGACATTGCAACATTATCTTAAGGAA 6754
1031 AsnLysAsp.....AsnLysTyrArgLys.. 1040
6755 CAGCTGACCAAAATACAGATGAGACTAAGAAATGAAGCTCAGAACTA 6804
1041 .....ThyLysAlaAspVal..... 1045
6805 TGAACCTCGCAAAAGATGATATCATGAAAGAAATCTCAGTCTGC 6854
1045 ..... 1045
6855 GTTAAATGCAAGACGCTCAGCAGAGAAAGATGATGTTGCAAGACT 6904
1046 .....GlnGlnLeuThrArgAspLleGlnSerLeuGlnGlnArg 1058
6905 ATGATATACCTGAGACAGAAACCAAGAAATACAGAGCTGATGAGAAA 6954
1058 GlnLeuSerLleGlySer.....LeuSerAlaLleGln 1070
6955 AATCTCGCTGTATTCAGACAAACACTTCTCAGAGCTCTCTCTA 7004
1070 LaaPleLysArgHisSerGln..... 1078
7005 GTGAGCTTCAAAAGCAACTGACACACAAACTTGCATGTTAAATA 7054
1079 LysGlnArgLeuAsnSerGlnPheAsnArgTPrGlnLysLeuSer.. 1094
7055 AAGAAATCTGTCATCAGCTCTCCAGATCTTGGACCTTGCAAAAC 7104
1094 ..... 1094
7105 TGACATGTTAAGCTAATATCACTCACTGACAGACCTTCTGAACAAATTA 7154
1095 .....ValTyrGlnSerAsnLleSerLysHisLysGlnLeuLysLeu 1109
7155 AGTTGTATACCGAAGCTGCTGCACTCAAGAAATCATAGCTTATCAA 7204
1110 SerGlnTyrLysAspLle.....GlnLysArgTyrThrAsnGlnPh 1123
7205 GATATATGGAAGAGACTGCTGAGCAAAAGAGCATGATGAGCTGCG 7254
1123 eLeuGlnLeuLysThrThrGlnMetAlaAsnLysAspLeuAspArgTyr 1140
7255 ACTCCAACTGCAAGTGTGGAGCAGCAGCAAGAAATGAGTGC..... 7297
1140 YrThrAlaLeuAspLysAlaLeuMetArgPheHisSerMetLysMetLys 1156
7298 .....GATTCTCAGCTGAGAACTCAAGTTCTGTGAATTTGAA 7336
1157 GlnLeuAsnLysLle 1162
7337 TTCTGATGATGATTACT 7354

```

seq_name: /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT.AA258980

seq_documentation_block:

ID AA258980 standard; cDNA; 7962 BP.

XX

AC AA258980;

XX

08-MAY-2000 (first entry)

DE Human cytoskeletal protein (HCYT) encoding cDNA (clone 1655208).
 XX
 XX Human cytoskeletal protein; HCYT; cell proliferation; immunological;
 KW reproductive; nervous disorder; cytostatic; immunosuppressive; anti-HIV;
 KW anti-diabetic; arteriosclerotic; dermatological; anti-inflammatory;
 KW anti-infectivity; vasotropic; cardiant; ss.
 XX
 XX Homo sapiens;
 OS
 XX
 XX Key Location/Qualifiers
 FT CDS 239..7567
 FT /tag= a

WO200006730-A2.

10-FEB-2000.

30-JUL-1999; 99WO-US17167.

31-JUL-1998; 98US-0155185.

04-AUG-1998; 98US-0160081.

19-AUG-1998; 98US-0155328.

(INCY-) INCYTE PHARM INC.

Bandman O, Tang YT, Yue H, Corley NC, Guegler KJ, Azimzal Y,
 PI Patterson C, Lal P, Baughn MR;

WPI: 2000-195297/17.

P-PSDB; AAY77575.

Human cytoskeletal proteins useful for diagnosing, treating preventing
 PT cell proliferative, immunological, reproductive, developmental and
 PS nervous disorders

claim 9; Page 80-83; 84pp; English.

XX The invention provides human cytoskeletal proteins (HCYT) and nucleic
 CC acids encoding the proteins. The HCYT polypeptides can be expressed by
 CC standard recombinant methodology. The HCYT polypeptides, modulators and
 CC antibodies are useful for treating or preventing a disorder associated
 CC with decreased and increased expression or activity of HCYT in mammals.
 CC The polypeptides are also useful for diagnosing HCYT activity disorders
 CC such as cell proliferative, immunological, reproductive, developmental
 CC and nervous disorders. Sequences AA258974-981 represent cDNA clones
 CC encoding the HCYT polypeptides.

Sequence 7962 BP; 2240 A; 1699 C; 2600 G; 1223 T; 0 other;

alignment_scores:

Quality: 449.00 Length: 1458
 Ratio: 0.591 Gaps: 60
 Percent Similarity: 52.126 Percent Identity: 20.645

alignment_block:

US-09-538-396-2 x AA258980 ..
 Align: seg 1/1 to: AA258980 from: 1 to: 7962

100 GlnLeuThrGlnLysAlaSerLysMetLysPheLysAlaLleGlnSerVa 116

2372 CAGTACATCAAGAGGCAAGCAAGCAAGAA...GAATGCTTGCACAGGCG 2418

116 LeuGlnThrLysAsnProHisThrGlnLysValCysLeuSerTyrA 133

2419 AGTCCAGAGAGAGAGGCCCTTAGTACAGAGAAAGCGCTTACAGGCTGC 2468

133 rGysAlaAspMetAspArgLys.....IleProAlaLeuMet 145

2469 GGCTGACAGCGCGTGGAGGCTGACCGCAGAGACCTGCTGACAACTACAG 2518


```
4184 ATGAACTACATGAACTATGACATCC.....TT 4212
      ::::: ::::: |||:::
658 uGlnAspAlaMetAsnLysArgAspGlnGlnLysArgLeuGlnAsnPhe 675
      |||::: ::::: |||::: |||::: |||::: |||:::
4213 ACAGAGTCGGCTG...CGAGACAGAGCTACAGCGAATGGA.....G 4253
      |||::: ::::: |||::: |||::: |||::: |||:::
675 LaaAsnGlyMetArgGlnMetLeu.....Ala 683
      |||::: |||::: |||::: |||::: |||::: |||:::
4254 CCCAGGCTGAGCGAGAGTACTTCAGGCGACCAAGAGAACTGACAGACC 4303
      |||::: |||::: |||::: |||::: |||::: |||:::
684 ProPheGlnHisLeu..... 688
      |||::: |||::: |||::: |||::: |||::: |||:::
4304 CAGGTGAACACCTGCAGACAGCTGTCTGTAAGAACCAAGGCTCAGGCAAG 4353
      |||::: |||::: |||::: |||::: |||::: |||:::
688 ..... 688
      |||::: |||::: |||::: |||::: |||::: |||:::
4354 TCGTGTGGCATCTGTGAAAGACCTGAGAACGGCTCGCTCAGACACTGA 4403
      |||::: |||::: |||::: |||::: |||::: |||:::
689 ..AlaArgLysAsnHisValLysProCysGlnArgAlaPheThrPro 704
      |||::: |||::: |||::: |||::: |||::: |||:::
4404 AGCTGAATAATGAGAAATGAGAGTAGAGTAGAGAGCCAGGCTCTG 4453
      |||::: |||::: |||::: |||::: |||::: |||:::
705 AspGlnGlnAspGlnPhe...ValLysGlnArgMetGlnAsnSerse 720
      |||::: |||::: |||::: |||::: |||::: |||:::
4454 CAAGAGCAGGGCGAACTGAAGGTGGCCCAAGGAGGCTCTGCAAGAGAA 4503
      |||::: |||::: |||::: |||::: |||::: |||:::
720 ThrAlaGlnArgSerLysAlaLeuAlaMetGlnSerSerAsnAlaGln 737
      |||::: |||::: |||::: |||::: |||::: |||:::
4504 TTGGCCCTCTGACCCAGACCTGAAAGAGAAAGAGAGGTGAGAG 4553
      |||::: |||::: |||::: |||::: |||::: |||:::
737 LalaLeu.....PheGlnGlnLeuAspLysLeuArgThrLysAsp 750
      |||::: |||::: |||::: |||::: |||::: |||:::
4554 CTCTGCGCGGACAAATCCAGAACTGAGAGACAAACGGAATGCGACAG 4603
      |||::: |||::: |||::: |||::: |||::: |||:::
751 AlaLysValLysLeuValGlnGlnThrLysProLeuAlaGlnLysAsnLe 767
      |||::: |||::: |||::: |||::: |||::: |||:::
4604 GCTGCTTGGAAATG.....CTGCTTGGACCTGAAAGAAAGAG 4661
      |||::: |||::: |||::: |||::: |||::: |||:::
767 uAsnGlnHisLeuAlaAspGlnSerGlnLysAlaGlnAlaPheAspArg 784
      |||::: |||::: |||::: |||::: |||::: |||:::
4662 GAACCAAGAGTAGATCTGCAGACAGAAACAGATTCAGAGCTAGAGAAAGT 4691
      |||::: |||::: |||::: |||::: |||::: |||:::
784 euLeuGlnValLeuAlaHisValGlnMet..... 793
      |||::: |||::: |||::: |||::: |||::: |||:::
4692 GTAGTGTGTTTAGACATCTGCCCATGGCCCTCCAGAGCGAGACAG 4741
      |||::: |||::: |||::: |||::: |||::: |||:::
794 .....AspArgAspAlaValGlnAlaLeuLeuGlnProThrAs 806
      |||::: |||::: |||::: |||::: |||::: |||:::
4742 AAGCTGACTGTGCAGAGCGAGACAGATCAGAGAGCTCGAAGAGATCGGGA 4791
      |||::: |||::: |||::: |||::: |||::: |||:::
806 PThrILAspArgHisValHisGlnIleGlnGlnLeuValLysGlnValG 823
      |||::: |||::: |||::: |||::: |||::: |||:::
4792 GACT...CAGAGAAACGCTCTTGAGCATCAGCTCTA.....GAACCTG 4832
      |||::: |||::: |||::: |||::: |||::: |||:::
823 LAspArgLeuLysValLeuAspSerSerGlnArgGlnValLysSerLeu 839
      |||::: |||::: |||::: |||::: |||::: |||:::
4833 AGAAGAAAGACCAATGATGATGCTCCAGAGAGGAGGAGGTTCAAGACCTG 4882
      |||::: |||::: |||::: |||::: |||::: |||:::
840 GlnLuu.....IleGlnGlnGlnLeuAs 847
      |||::: |||::: |||::: |||::: |||::: |||:::
4883 AAAAAGCAGTTGGTACTCTGGAATGCTGCGCTCGAATCTGGAGAAA 4932
      |||::: |||::: |||::: |||::: |||::: |||:::
847 nPhe.....LeuGlnArgT 852
      |||::: |||::: |||::: |||::: |||::: |||:::
4933 CCATCACAAGATGAGAGTCCCAAGAAAACATGATCAGAGAGCTGAGGGCC 4992
      |||::: |||::: |||::: |||::: |||::: |||:::
852 hrArgAspThrLeuIleValGlnValAsp.....AspLeuArgT 864
      |||::: |||::: |||::: |||::: |||::: |||:::
4983 AGAGGAAACCCAGAGAGTGGCTTTGACCCACCTTACGCTGACCTGAGAA 5032
      |||::: |||::: |||::: |||::: |||::: |||:::
865 AspGlnHisArgMetLeuAsn.....G 872
      |||::: |||::: |||::: |||::: |||::: |||:::
```

```

1149 gphetisermetylmetgluile.....Asn 1160
1150 .....
5933 CCTCAGACAGCAGTGTGCTGACACAGACAGACATGAGTGGAGACCA 6002
1160 ysilellelysgluLeuTrpGlnGlnThrTyf.....Arg 1171
1161 .....
6003 GGGCCCTCAGACAGCAGTGTGCTGACAGCCAGCAGTGTCTCAAGAACGG 6052
1172 glyGlnaspIleAspTrpIleSerIleAsnSerAspSer..... 1184
1173 .....
6053 GACCAAGAGCTGGAAGCTCTGCGGAGCAAGTCACTCTCCCGGACATCA 6102
1185 ..... GluGlyAlaGlyThrArgSerTyfSerTyfArgValValMetGlnT 1200
1186 .....
6103 GGAGAGAGCTGCGCGCGCGCGGCTGAGGCTGTGACAGAGGCGCTT... 6148
1200 hrgIAspAlaGluLeuGluMetArgGlyArgCysSerAlaGlyGln 1216
1201 .....
6149 ..GGCAAGGCTCATGCTGCGCGCGGAGAA.....GAGCAGCAT 6187
1217 ValLeuAlaSerLeuIleIleArgLeuAlaLeu...AlaGluThrPheCy 1232
1218 .....
6188 CTCTCGAGCAGGACAGAAATTGAGCCGAGCTGTGAGGCGCAGCAGTGCAC 6237
1232 sLeuAsnGlyGlyIleLeuAlaLeuAspGluProThrThrAsnLeuAsp 1249
1233 .....
6238 CCTGCAACCTCTGCGGAGTGTGCGGAGCAGCAGTGTGCGGAGTGTGAG 6286
1249 LyrProAsnAlaGluSerLeuAlaAlaLeuLeuArgIleMetGlnAla 1265
1250 .....
6287 .....GAGGCTGTGAGGATACAGAGAGT 6310
1266 Arg.....LysGlyGlnGluAsnPheGlnLeuIle.. 1275
1267 .....
6311 GAGATCCAGACAGCAGATCTCCGATACAGAGATGTGACAGCAGCTGCA 6360
1276 .....ValIleThrHisAspGluArgPheAlaHisLeuIleGlyGlnA 1290
1291 .....
6361 GCAGCAGCTTGTCCGAGAGGATGAGAGCTGAGACTGACAGAGAACGGG 6410
1290 rGlnLeuAlaGluLeuTyfTyfArgValSerLysAspGluAsnGlnHis 1306
1291 .....
6411 AGCAGCTCTGTGAGAACTCTGCGCCAGAGGCTCCAGAGAAATATG... 6457
1307 SerIleIleGluSerGlnGluIle 1314
1315 .....
6458 ...ATCCAAGAGACAGCAGATCTG 6478

seq_name: /SIDS1/gcgdata/geneseq/geneseq/NA1996.DAT:AAAT34291
seq_documentation_block:
ID AAT34291 standard; cDNA to mRNA; 6175 BP.
XX
AC AAT34291;
XX
DT 25-MAR-1997 (first entry)
XX
DE Coding sequence for smooth muscle myosin heavy chain SM1.
XX
KW Smooth muscle; myosin heavy chain; SM1 isoform; rabbit; arteriosclerosis;
KW gene therapy; mouse; SM2 isoform; retrovirus; adenovirus; restenosis;
KW associated adenovirus; coronary artery catheterisation; sclerotic artery;
KW ss.
XX
OS Mus musculus.
XX
PH Key Location/Qualifiers
FT CDS 105..6023
FT /tag= a
FT /product= SM1
XX
PN W09623069-A1.

```

```

XX
PD 01-AUG-1996.
XX
PE 25-JAN-1996; 96WO-JP00134.
XX
PR 25-JAN-1995; 95JP-0010085.
XX
PA (OSAP) OSAKA PREFECTURAL GOVERNMENT.
XX (VSS-) VESSEL RES LAB CO LTD.
XX Arakawa E, Hasegawa K, Ishiyama H, Matsuda Y, Oda S;
XX Sugawara M, Takahashi K;
XX WPI: 1996-362693/36.
XX P-PSDB: AAM00024.
XX
PT DNA encoding smooth muscle myosin heavy chain SM1 isoform - used in
PT a vector for gene therapy for reduction of re-stenosis following
PT coronary artery catheterisation
XX
PS Claim 5; Page 14-27; 42pp; Japanese.
XX
CC This sequence represents the coding sequence for the smooth muscle myosin
CC heavy chain SM1 isoform protein. This SM1 coding sequence was isolated
CC from a mouse embryo cDNA library using probes based on the rabbit smooth
CC muscle myosin heavy chain SM2 isoform. This sequence is included in the
CC gene therapy vector of the invention. The vector of the invention
CC consists of this sequence inserted into a retrovirus, adenovirus,
CC associated adenovirus or animal-expression plasmid vector (such as pCMX2
CC or pAGE208). The vector can be used in the gene therapy treatment of
CC arteriosclerosis, particularly for the reduction of restenosis occurring
CC following coronary artery catheterisation for widening of sclerotic
CC arteries.
XX
SQ Sequence 6175 BP; 1846 A; 1455 C; 1056 G; 1018 T; 0 other;

alignment_scores:
Quality: 446.00 Length: 1494
Ratio: 0.599 Gaps: 71
Percent Similarity: 49.799 Percent Identity: 20.750

alignment_block:
US-09-538-396-2 x AAT34291 ..

Align seg 1/1 to: AAT34291 from: 1 to: 6175

13 ArgSerPheAspProAspAsnLysAsnValIleThr..... 24
1884 AAGAACATGACCCCGCTAAATGACATGTGACATCTCTCAATGCCCTC 1933
25 .....PhePheLysProLeuThrLeuIleValG 34
1934 CTTGACAGATTGTGGCTGACCTGTGAGAGATGTGAGACCCGATTTGG 1983
34 LyrProAsnGlyAlaGlyLysThrThrIleIleGluCysLeuLysLeuSer 50
1984 GGCTGAGCCAGATGGCCAGATGACT..... 2009
51 CysThrGlyGluLeuProProAsnSerArgSerGlyHisThrPheValHI 67
2010 ...GAGAGCTCACTGCGCAGTGTCTCAAGACC..... 2039
67 sAspProLysValAlaGlyGluThrGluThrLysGlyGlnIleLysLeuA 84
2040 .....AAAAAGGCGATGTCGCGACCGTGGAGACGCTC..... 2072
84 rPheLysThrAlaAlaGlyLysAspValValCysIleLeuArgSerPheGln 100
2073 ..TACAAAGACAGCTTGGGAAACTGATGTGCTACACTGTGCGAATACCA 2120
101 LeuThr.....GlnLysAlaSerLys 107

```

2121 GCTAATTGCGCGCTGCATCATCCCAACATGAGAGAGCTGCGCAA 2170
 107 smetglu...Phelysalaileglu.....SerValleug 118
 2171 GCTGATGATCTTCTAGTGTGAGAACAGCTGCGCTGCACAGGTGTGG 2220
 118 lmrthleasn.....Prohis..... 123
 2221 AAGCATCCGATCTGCGCTGAGGGCTTCCCAACAGAGATTGCTTCCA 2270
 123 123
 2271 GAGTTCGGGCAACGCTATGAGATCCTGGCAGCAACGCCATCCCAAG 2320
 124Thrglylulysvalys...LeuserTyrArgCysAlaaspm 137
 2321 CTTCATGATGAGAAAGCAAGCCTCATCTTCATGATCAAGCCCTCGAAC 2370
 137 etaspatggluileproalaleu...Metglyvalserlysalaval 151
 2371 TTGAC.....CCTAACCCTGTACAGAGATTGGCAGAGCAAAATCTTC 2411
 152 leugluasnValillephevalhisgluasnpsluserasntrproleugl 168
 2412 TTCGGAAGGGGGTCTGCGCCCACTAGAGAGAGACA..... 2450
 168 nasproserthreuleuylslyspheaspsapillepheseralathra 185
 2451GACGTGAAATTAATGATGATCATGCGCTTCC 2484
 185 rg.....TyrThrlys..... 188
 2485 AGGCAATGTGTGCGCTACCTGCGCAAGAAAGCCTTCACCAAGAGCGAG 2534
 189AlaleuGlulValillelys..... 194
 2535 CAACAGCTGACAGCATGAGGTATGATCCAGAGAACTGCGCTCACT 2584
 195 LysleuHisLysaspGlnMetGlnLulLysThrPheArgLeuLysL 211
 2585 TAAAGCTCGCAACTGGCAATGTGGCGCTCTCC.....AAG 2625
 211 euGlulasnleuGlnThrValLysaspGlnAlaHisLysLeuArg..... 225
 2626 TAAAGCCATGCTCCAGGTGACACGGCAGAGAGAGATGACAGCCCAAG 2675
 226GluasnillealaglaspGlnGlnLysSeraspAlase 238
 2676 GAGGAGGATGACAAAGATCAGAGCGCGCAGCAAGAGCGAGACTGA 2725
 238 rlyserGlnMetGlnLulLysLysIleCysglYThrleu... 253
 2726 GTTGAAGAGGTGAGCAGAGACACTAGCTGCTGAGGAGAACTC 2775
 254ArgGlulileuGlnMetGlnThrSerleu...AspGlnleuArg 267
 2776 TCGTCGACAGAGAGTTGACAGCAGACAGAGCTGATGCTGAGCTGAG 2825
 268 ArgleuGlnGlnLulLysaspilleysalathrgluArgserThrleu 284
 2826 GAGATGCGGGTCCGTTGGCAGCCAGAAAGCAGAACTGAGAGAGATCT 2875
 284 uThrglnHisglulysLysleuAlaLysSerGlnLulLysanglu... 299
 2876 A.....CATGAGATGAGAGCGCGCTGAGGAGAAAGAGACCGGC 2916
 300Asp 300
 2917 GCCACCACTACAGGCTGAGAGAAAGATGCTACAGCAGATGCTAAC 2966
 301 ThrAspGlnLulLysMetGlnLulLysThrPheGlnLulLysPheGlnLul 314
 2967 CTGGAGAGCAACTGAGAGAGAGAGCGCGCCAGACAGAAACTACACT 3016
 314 uArgillealaleuLulLysLysIleSerLysleuValArgasp... 329
 3017 AGAGAAAGTCAAGCGCTGAGCGCAAGATCAAGAACTGAGAGATGACTCT 3066
 330MetAspAspGlnLulLysSerSerValleuSerLysGlnasn 344
 3067 TGTATGATGATGATCAG.....AACAGTAAACTTTCAAAAGAACGA 3107
 345 SerGlnleuThrHisglulileglyLysleuGlnAlaGlnAlaAspAla 361
 3108 AAACCTCTTGAAGAGAGGCTCAGCAGCTTG.....ACAAACCA 3145
 361 sleuthMetLysHisgluArgaspsersapilleysasnIleCysThrL 378
 3146 CTTAGCAGAAAGAGAAAGGCTAAAGCTCAAGAAAGCTGAGAGCA 3195
 378 yshIsasnleuGlyProValProGlnHisProPheThrAsnAspVala 394
 3196 AGCAT.....GAGTCTATGATCTCAGAGCTGAG 3224
 395 MetasnleuThrAsnArgilleysalathrgleuser..... 406
 3225 GTGAGGCTGAGAAAGAGAGAGAGAGCGCCGAGAGCTGAGAAACTCA 3274
 407SerleuGluasnAspLeuLysLysSerasnGlnuAsp. 421
 3275 GAGGAACTGAGAGGGTGTGCTGACATGACTTCATGAGAGATCGTGCT 3324
 422GlnleuAspVal.....LeuTrpLysHisThrleu 431
 3325 TGCAAGCCCAATGTCAGAGCTCAAGATGCAAGCTGCGCAAGAAAGGAA 3374
 432 LysIleasnAlaArgTyrSerGlnValAspGlnIleGln..... 445
 3375 GAGCTACAGCAGCTCTAGCCAGCTTGTATGAGAGATCGCCCAAGAAA 3424
 446SerLysIleGlnSerMetSerGlyLleu..... 455
 3425 CAATGCCCTAAAGATTTGCGAGCTAGAGGGCCATATCTCAGACCTAC 3474
 456ArgArg 457
 3475 AAGAGACTAGACTCAGAGCGGCTGCCAGGAACAAGCCGAGAGAGAG 3524
 458 ArgLysaspLysGlnLysGlnArgAspAlaAlaGlnValLulLysSerly 474
 3525 AAGCAGAGACTGGGGAGAGAGCTGAGGACACTCAAGAGAGCTGAGAGA 3574
 474 sphenasnleuserArgilleaspslulArgGlu.....ArgHisM 487
 3575 TACGCTGACAGCAGCTACCCAGCAGAGAGCTCAGACCAAGAGAGAAC 3624
 487 etGlnIleGlnValGlnArgLysThrleuAlaLysGlnLulLysArgspTyr 503
 3625 AGGAGCTACAGCTGCTGAAGAGCGCTGAGTGAAGAGAGCGGTCCAT 3674
 504 AspSerIleIleSer.....GlnLysArgThrGlu...ValTyrSe 516
 3675 GAGGCCAGGTCACAGAGATGAGCAGCAGACACAGCAGCAGCTGAGAGA 3724
 516 rleuGlnLulLysIleLysValleuLysArgLulLysaspIleleasna 533
 3725 ACTCAAGAGCAGCTGAGCAGCTCAAAAGGGCCAAAGCAAACTGGAGA 3774
 533 rg.....AsnAlaAsp.....GlnuArg 538
 3775 AAAGCAACAGACACTGAGAGAGAGAAAGCGGACCTGCTGAGAGCTG 3824
 539 ValLysleuGlyLulLysLysAspAlaLysleuGlnLulLysSerLysAspLysle 555
 3825 CGTGCTCTGGCCAGCAGAGAGAGAGGTGGAACACAGAAAGAGAGAGCT 3874

5510 GGAGTAGAAGTGTCTCAAGCAAGCTCAAGTCTGCGGCC 5559
 1068 legluAlaAspLeuLysArgHisSerGlnGluLysGluArgLeuAnSer 1084
 5560 TGGAGGCAAGATT.....GCACAGCTGGAGGAGCAGGTTGAACAG 5600
 1085 GluPheAsnArgTrpGlnGlnGlyThrLeuSerValTyrGlnSerAnIleSe 1101
 5601 GAGGCCAGAGAAACAGGCGGCCCAAGTCCGTGAAGCAAGAACAGCA 5650
 1101 rLysHisLysGln.....GluLeuLysLeuSer. 1110
 5651 GAGCTAAAGAGAGCTCTGTGAGGTGAGATGAGCCGAAGATGGCAG 5700
 1111 ..GlnTyrLysAspIleGluLysArgTyrThrAsnGlnPheLeuGlnLeu 1126
 5701 AGCAGTACAAAGAGCAGCAGCAAGAAAGAACACCAAGGTCAAGCAGCTG 5750
 1127 Lys.....ThrThrGluMetAlaAsnLysAs 1135
 5751 AAGAGCAGCTGGAAGAGCAGAGAGAGTCCAGTGCATCAACGCCAA 5800
 1135 PLeuAspArgTyrTyrThrAlaLeuAspLysAlaLeuMetArgPheHis 1152
 5801 CCGCAGGAACCTGCAGCGGAGCTAGTAGAGCC.....A 5835
 1152 erMetLysMetGluGluIleAsnLysIleIleLysGluLeuTrpGlnGln 1168
 5836 CAGAGAGCAATGAGGCCATGGCGGTGAGTGAAGCCCTCAAGAGCAAA 5885
 1169 ThrTyrArgLysGlnAspLysPheLysSerIleAsnSerAspSerGln 1185
 5886 CTAGAGGAGAAACGAGGCTCATTTGTT.....CCTCCAG 5923
 1185 uGlyAlaGlyThrArgSerTyrArgValAlaMetGlnThrGlyA 1202
 5924 AAGGCTGGGGGCGCT.....AAGTATATGAAACACAGATG 5961
 1202 sPaLaGluLeuGluMetArgLysArgCysSer 1212
 5962 GTTCTGAAGAAAGATGGACGCTGGGAGCTCA 5993

seq_name: /stds1/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV21518

seq_documentation_block:
 ID AAV21518 standard; cDNA; 5661 BP.

AAV21518;
 17-AUG-1998 (first entry)

Rattus norvegicus mutant alpha-myosin heavy chain transgene.
 Actin binding domain; alpha myosin heavy chain; bridging; mouse;
 transgenic mammal; congestive heart failure; study; treatment; diet;
 exercise; effects; identification; hypertrophic cardiomyopathy;
 dilated or hypertrophic cardiomyopathy; acute aortic regurgitation;
 tricuspid stenosis; constrictive pericarditis; hypertension;
 acute infective endocarditis; ischaemic heart disease;
 primary myocardial disease; valvular disease; pericardial disease;
 hyperthyroidism; anaemia; arteriovenous fistula; beriberi;
 Paget's disease; transgene; ss.

Rattus norvegicus.
 Synthetic.

Key Location/Qualifiers
 CDS 1..5661

/tag "mutant alpha-myosin heavy chain"

MO9813476-A1.

PD 02-APR-1998.
 XX 26-SEP-1997; 97MO-US17296.
 XX 26-SEP-1996; 96US-0026742.
 XX (LEIN/) LEINWAND LA.
 XX Vikstrom KL;
 XX WPI: 1998-230690/20.
 DR P-PSDB: AAM54241.
 XX
 PT Transgenic mouse models for congestive heart failure and
 PT hypertrophic cardiomyopathy - used to study molecular and cellular
 PT events; identify potential therapeutic agents; assess effects of
 PT diet etc.
 XX
 PS Claim 39; Pages 43-53; 75pp; English.

The sequence is that of a mutant rat alpha-myosin heavy chain
 gene which was used in the development of transgenic mammals,
 specifically mice. They can be used as a model for studying congestive
 heart failure (CHF) or hypertrophic cardiomyopathy. Such animals
 are used to study molecular and cellular events associated with
 CHF; to identify compounds for treating CHF; and in evaluating
 effects of diet and exercise on CHF. Conditions associated with CHF
 that can be evaluated this way are dilated or hypertrophic
 cardiomyopathy; acute aortic regurgitation; tricuspid stenosis;
 congestive pericarditis; acute infective endocarditis; ischaemic
 heart disease; hypertension; primary myocardial disease; valvular
 or pericardial disease; hyperthyroidism; anaemia; arteriovenous
 fistula; beriberi and Paget's disease.

Sequence 5661 BP; 1572 A; 1446 C; 1797 G; 846 T; 0 other;

alignment_scores:

Quality: 445.00 Length: 1403
 Ratio: 0.614 Gaps: 59
 Percent Similarity: 51.675 Percent Identity: 19.743

alignment_block:
 US-09-538-396-2 x AAV21518

Align seg 1/1 to: AAV21518 from: 1 to: 5661

98 SerPheGlnLeuThrGlnLysAlaSerLysMetGluPheLysAlaIleG1 114
 1774 TCCTCCAGACAGTGTCTGCTCCACCGGAAATCTGAACAAGCTGAT 1823
 114 uSerValLeuGlnThrIleAsnProHisThrGlyGluLysValCysLeus 131
 1824 GACAAACCTGAGCACCCATCCTCAC.....T 1852
 131 erTyrArgCys..AlaAspMetAspArgGluIleProAlaLeuMetGly 146
 1853 TTGTGCCCTCATCATCCCAATGACGCGAAGCTCA..... 1890
 147 ValSerLysAlaValLeuGluAsnValIlePheValHisGlnAspGluSe 163
 1891GGGGTATGAGCAACACCCCTGTCATGCACACAGTCGATG 1931
 163 rAsnTrpProLeuGlnAspProSerThrLeuLysLysLysPheAsp... 178
 1932 CAACGAGTGTGAGGAGGTATCCGCATGTGTAGAAAGGCTTCCCAACC 1981
 178 178
 1982 GCATTCTTATGGGACTCCGGCAGAGGTATGATTCGATCCGAACCCAGCA 2031
 178 178

2032 GCCATCCCTGAGGGCCCAATTCATTGATACGGGAAAGGGCTGAGAAGCT 2081
179 Aspie..... PheSerAlaThr 185
2082 GCTGGGCTCCCTGGACATTGACCAACACAGTACAAAGTTGGCCACACCA 2131
185 rg... TyrThrLysAla..... LeuGluValIleLysLysLeuHISLys 198
2132 AGGTGTTCTTCAGAGCGGGCTGCTGGGCTGCTGGAGGAGATGCGAGAT 2181
199 AspGlnMetGlnIleLysThr..... 206
2182 GAGAGCGCTAGCCGCATCATCACCAATTCACAGCTCAGCCGAGGCCA 2231
207 ... PheArgLeuLysLeuGlnLysLeuGlnThrValLysAspGlnAlaH 222
2232 GCTCAAGCCGATAGTTGAGTACAGAAAGTGTGAGCCGAGGAGCGCCCTGC 2281
222 LeuLysLeuArgLysAsnIleAlaGlnAspGlnIleLysSerAspAlaSer 238
2282 TGGTTATCCAGTGAACATC..... CGCGCTTCATGGGG 2316
239 LysSerGlnMetGlnIleLysLys..... G1 247
2317 GTCAGAAATGGCCGCTGATGAAGCTCTACTTCAGATCAGCCGCTGCT 2366
247 uLysIleCysGlyThrGluArgGluIleLeuGlnMetGluThrSerLeuA 264
2367 GAAGAGCCGAGACAGAGAGAGATGGCCACACATGAAA..... G 2407
264 spGluLeuArgArgLeuGlnGlyGlnIleAspIleLysAlaThrGluArg 280
2408 AGGAGTTCGGGAGTCAACAGATGCCTAGAGAGTCTGAGGCTCCGCGC 2457
281 SerThrLeuLeuThrGlnGlnHISGlyLysLeuAlaLysSerGluG1 297
2458 AAGGAGCTG..... GAGGAGAAAGATGCTGCCCTGCTGCAGGA 2495
297 uAsnGluAspThrAspGlnGluLeuMetGluThrGlnThrLysPhe... 312
2496 GAAGATGCTCAGCTCCAGTCCAGTCCGAGCGGAAACAAACCTGGCAG 2545
313 GlnGluArgIle..... AlaLeuGluGln 320
2546 ATGCCGAGAGGCGCTGCGACAGCTGATCAGAACAAAGATCCAGCTGGAG 2595
321 ThrLysIleSerLysLeuValArgAspMetAspAspGluAlaSerTyrSe 337
2596 GCCAGAGTCAAGAGATGACCCAGAGAGCTGAGAGCAGAGAGAGATGAA 2645
337 rSerValLeuSerLysGlnAsnSerGluLeuThrHISGluIleGlyLysL 354
2646 CCCGAGCTCAGCGCCAAAGACGCAAGCTGGAAGCAGAGTCTCAGAGC 2695
354 euGlnAlaGlnAla... AspAlaHISLeuThrMet..... 364
2696 TCAGAAATAATATCGATGACTGAGCTGACCTGGCCAAAGTGAGAAAG 2745
365 ... LysHISGluArgAspSerAspIleLysAsnIleCysThrLysHISAs 380
2746 GAAAGGACGACGACAGAGAACAAAGTTAAAAACCTG... ACAAGAGAGAT 2792
380 nLeuGlyProValProGlnHISProPheThrAsnAspValAlaMetAsnL 397
2793 GCGCGGG..... CTGAGCAGATCATTTGCCAAGC 2821
397 euThrAsnArgIleLysAlaArgLeuSerSerLeuGlnLysAsnSerLeu 413
2822 TCACCAAGGAGAGAAAGCT... CTTCAGAGAGCCACACAGCAAGCCCTA 2868
414 AspLysLysLysSerAsnGlnLysAspGlnLysAspValIleuThrLysHIS 430
2869 GATGACCTTCAGAGCTGAGAGACAAAGTCAACACACTGACCAAGTCTAA 2918
430 IleuLysIleAsnAlaArgTyrSerGluValAspGlnGlnIleGlnSer 447
2919 AGTCAGCTGAGACACAGGTGATGATCTGGAGGATCTCCGAGACAG 2968
447 ysiIleGlnSerMetSerGlyIleLeuArgArgArgLysAspLysGlyLys 463
2969 AG..... AACAAAGTGGCATGACCTGGAGCCA 2997
464 GluArgAspAlaAlaGlnValGluLeuSerLysPheAsnLeuSerArgI1 480
2998 GCAAGCGGAGAGCTGAGGCTGACCTGAGCTGAGCTGACCCAGAGAGACTCAT 3047
480 eaSpGluArgGluArgHISMetGlnIleGluValGluAlaGlySThrLeuA 497
3048 GGAAGCTGAGAACAGACAGCTTCAGCTGAGGAGAAAGCTCAAG..... 3090
497 LeuGlnGluArgAspTyrAspSerIleIleSerGlnLysArgThrGln 513
3091 AAGAAAGATTGAC..... ATCAGTCAAGCAAGACATTA 3126
514 ValTyrSerLeuGlnGlnLysIleLysValLeuLeuArgLysAspI1 530
3127 ATAGAGAGAC... GAGCAGGCCCTGGCCCTTCAGCTGACAGAGAACTGAA 3173
530 eileAsnArgAsnAlaAspGluArgValLysLeuGlyLeuLysLysAspA 547
3174 GGAAGAACAGAGCAGCATGAGAGCTGAGGAGGAGAGCTAGAGCCGAGAC 3223
547 LeuGlnLysSerLys... AspLysLeuAsnGlnIleValAsnGlnHIS 562
3224 GCACAGCCCGGCGCAAGGTGAGAGACCTGCTGCACAACTGACCCGCGAG 3273
563 LysAspLysIleLysLysValLeuArg..... GlyArgAsnProh 576
3274 CTGAGAGAGATCATGAGAGCTAGAGAGAAAGCGGTGGGCCCACTGCT 3323
576 eGluLysAspMetLysGlnIleAsnGlnAlaPheThrProValAspL 593
3324 GCAGATAGAGATGACAAAGAGCGGAGCGCGAGCTCCAGAAAGATCGGC 3373
593 yAGIuTyrAsnGln..... Leu 598
3374 GGGACCTGAGAGAGCCAGCTGACAGATGAGGCCACAGCCGCGGCTG 3423
599 ArgSerLysSerGlnLysAlaGlnGlnLysLysPheThrGlnSerLys 615
3424 CGCAAGAGAGCAGCAGACAGCGTGGCCGAGCTGGGAGACAGATACAA 3473
615 sValThrAspAlaArgGlnGlnIleuThrLysLeuArgArgAspMetAspA 632
3474 TCTACAGCGGCTGAGACAGAACTGGAGAAAGAGAGACGAGTTCAAC 3523
632 LeuLysArgArgPheLeuAspSerLysLeuGlnSerIleLeuGlnIleSer 648
3524 TGGAGCTGATGACGTACCTCTCACATGAGAGATCATCAAGGCCAAG 3573
649 AlaAsnValAspMetPheProLysValLeuGlnAspAlaMetAsnLys... 664
3574 GCAAAGCTGAGAAAGTGCCTCCGAGACACTGAGAGACAGGCAATTAATA 3623
665 ArgAspGlnGlnLysArgLeuGlnLysAsnPheAlaAsnG 677
3624 CCGGCTGAAAGTTGGAGAAAGCCAGCGCTCCCTCAATGACTTCACACAC 3673
677 LysMetArgGlnMetLeuAlaProPheGlnHISLeuAlaArgLysAsnHIS 693
3674 AGCGAGCCAAAGCTGCAGACAGAGATGCGAGTGGCTGGGAACTGGAA 3723
694 ValCysProCysCys..... GluArgAlaPheH 703
3724 GAAAGAGAGCACTGATTTGGCAGCTGACCCGCGGGAACCTCTCCATAC 3773

703 rProaspGluGluAspGluPheValIysGlnArgMetGlnAsnSers 720
704
3774 CCACGACATGAGAGC.....CTCAAGAGCAGCTGAGAGAGAG 3814
720 eThralaGluArgSerIysAlaLeuAlaMetGluSerSerAsnAlaGlu 736
721
3815 GCAGGCCCAAGATGCTTGCCACGCACTGCGTCAAGCCCGGCAAGAC 3864
737 AlaLeuPheGlnGlnLeuAspIysLeuArgThrIleTyrAspAlaTyr 753
738
3865 TGC.....GACCTGCTCGGGACAGATACGAAGAAGAAAT 3899
753 LysLeuValGluGluThrIleProLeuAlaGluLysAsnLeuAsnGln 770
754
3900 GGAGGCCAAGCCGAGCTGAGCTGCTCTCTCCAG...GCCACTCAG 3946
770 LsLeuAla.....AspGluSerGlnLysAla 778
771
3947 AGGTGCCCACTGAGAGCAAGTATGAGAGCGGCCATACAGAGAGC 3996
779 GlnAlaPheAspAspLeuGluLysValLeuAla...HisValGlnMetAs 794
780
3997 GAGAGAGCTGAGAGAGCAAGAGAGAGCTGCTCAGAGCTTCAAGATGC 4046
794 PArgAspAlaValAlaLeuLeuGlnProThrAspThrIleAspArg 811
795
4047 TGAGAGGAGCTGAGAGCGCTCAACGCCAAGTCTCTCCCTGAGAGAGA 4096
811 LsValHisGluIleGlnLeuValIysGlnValGluAspLeu..... 825
4097 CCAGACACAGGCTGAG.....ACGAGATCGAGAGACCTGATGCTG 4137
826GluTyrAlaLeuAspSerSerGlyArg 834
4138 GATGTGAGCGCTCCAAATGGCGCGCGCAGCCCTGACAAAGAGAGAG 4187
834 GgLyValIysSerLeu.....GluGluIleG 843
4188 GAACCTCGACAAAGATCCTGCTGAGTGCAGACAGAAATGAGAGAGTCC 4237
843 LInleuGluLeuAsnPheLeuGlnArgThrArgAspThrLeuIleValGlu 859
844
4238 AGTCAGAGCTGAGAGCTTCCAGAAAGAGAGCGCGCTCCCTGAGCAGAG 4287
860 ValAspAspLeuArgAspGlnHisArgMetLeuAsnGluAspMetSer 876
861
4288 CTCCTCAAGCTCAAGAAATGCTATGAGAGAGTCTGAGCAGCAGTGAAG 4337
876 rAlaGlnValArgTyrHisAsnAlaArgGluLysValIysAlaSers 893
877
4338 CTCACAGCGGAGAAACAAGACCTCCAGAGAGAG.....ATCTCAG 4378
893 eIleLeuGluArgPheGlnLysSerGluGluValLeuValLeuAla 909
894
4379 ACCTGACTGAACGCTGGAGAGAGGGGTAAATGTGCACAGAGCTGAG 4428
910 GluGluLysGluGlnLeuIleValGluLys.....LysLeuLe 922
911
4429 AAGATCCCAACACAGCTGAGTGAAGAGCTGGAAGCTGACAGCTCCT 4478
922 uGluLysLeuAspProLeuSerLysGluLysGluSerLeuLeuGln 938
923
4479 GGAGGAGGCTGAGCGCTCCCTGAGCAGATGAGAGGCAAGATCCCTCAG 4528
939GluTyrAsnAlaLeuLysGlnLysLeuAspGluGluTyrHis 952
940
4529 CCCAGCTGAGTTCACACAGATCAAGGCAAGAGATGAA.....AGG 4569
953 GlnLeuAlaGluArgLysArgGluPheGlnGlnLysLeuAspAlaLeuGlu 969
954
4570 AAGCTGGCAGAAAGAGAGAGATGAGAGCGCAAGCGCAACCACT 4619
969 yArgLeuAsnMetLysIleLysGlyTyrLeuAspSerLys.....L 983

4620 GCGGCTGTGAGACTCCCTACAGACCTCCCTGATGCCGAGACAGCAGCC 4669
983 yAsnGluLysLeu.....LysGluLeuGlnGlyArgHisValLeu 996
4670 GCAAGAGCGCTGCGGCTGAGAGAGAGATGAGAGCGAGC..... 4710
997 CysHisSerGlnLeuGlnSerCysMetAlaLysGlnGlnArgIleSerAl 1013
4711CTCATGATGAGAGATCCACCTCAGTCAAGGCCAATGAAATAGCTC 4757
1013 aGluLeuAsnLysSerLysGluLeuLeuGlnGlyGlnGlnLysAla 1030
4758 AGAG.....GCCAGAGAGCACTGAGAGAGAGCCCAAGCCCACTGAG. 4800
1030 rGAsnIleAspAsnLeuLysTyrArgLysTyrLysAlaAspValGlu 1046
4801GACACCAAGCTCCACCTGAT 4821
1047 GlnLeuThrArgAspIleGluSerLeuGluGluArgLeuLeuSerIleGlu 1063
4822 GAGCAGCTCCTGCCAATGAGAGCTGAGAGAGAGCAATCGCCATCTG. 4869
1063 ySerLeuSerAlaIleGluAlaAspLeuLysArgHisSerGlnLysG 1080
4870G 4870
1080 LysArgLeuAsnSerGluPheAsnArgTyrGlnGlyThrLeuSerValTyr 1096
4871 AGCGCGCAACACCTGCTG.....CAGCGGAGCTGAGAGAGCTG 4911
1097 GlnSerAsnIleSerLysHisLysGlnLysLeuLysLeuSerGlnTyrLys 1113
4912 CGGGCGCTGTGAGAGACAGACAGCGCTGGAAGCTGAGAGAGAGAG 4961
1113 sAspIleGluLysArgTyrThrAsnGlnPheLeuGlnLeuLysThrThrG 1130
4962 GCTGATCGAGACAGCAGAGCGCGGTGACACTGCTGACATCCAGAAACCA 5011
1130 LMetAlaAsnLys.....AspLeuAspArgTyrTyr 1140
5012 GCTCATCAACCAAGAGAGAGATGATGATGACACTTCTCCAGCTCCAG 5061
1141 ThrAlaLeuAspLysAlaLeuMetArgPheHisSerMetLysMetGlnGlu 1157
5062 ACAAGGTGAGAGCGCGTGCAGAGGTGTAGAGAGCAGAGAG...GAGAA 5108
1157 uIleAsnLysIleLeu.....LysGluLeuTyrG 1167
5109 GCCCAAGAGCCCATACAGATGCCCGCATGATGCCGAGAGAGCTGAGAGA 5158
1167 GlnGlnThrTyrArgGlyGlnAspIleAspTyrIleSerIleAsnSerAsp 1183
5159 AGGAGCAGACACAGCGCCCACTGAGCGCATGAGAAAGAACATGAG 5208
1184 SerGluGlyAlaGlyThrArgSerTyrSerTyrArgValValMetGlnTh 1200
5209 CAGAGC.....ATCAGAGACTGAGAGACCGCGTG..... 5238
1200 rGlyAspAlaGlu...LeuGluMetArgGlyArgCysSerAlaGlyGlnL 1216
5239 GACAGGCAAGACGATCGCCCTCAAGGT.....GGCAAGA 5275
1216 yValLeuAlaSerLeuIleLeuArgLeuAlaLeuAlaGluThrPheCys 1232
5276 AGCAGCTCAGAAACTGAGGCCCG..... 5301
1233 LeuAsnCysGlyIleLeuAlaLeuAspGluProThrThrAsnLeuAspGlu 1249
5302GTCGGGAGCTGAGAGATGAGCTGAGAGCTGAGAGAGAA 5339
1249 yProAsnAlaGluSerLeuAlaAlaLeuLeuArgIleMetGluAla 1266

```
5340 GCGCATGGGAGTGGTGAAGGC.....ATGAGAGAGAGCGCGC 5383
1266 rglYsclYnglnuAnpneGlnLeuIleValIleThrIAspGlu... 1281
5384 GCATCAAGAGCTCACCTACAG.....ACAGAGAGAGACAG 5421
1282 .....ArpHeAlaHisLeuIleGlyGlnArgGlnLeuIleAgi 1294
5422 AAGAACTGTGCTGGCTGCAGACGCTGTGACAACTGCTGTAAGGT 5471
1294 ulYsTYrTYrArgValSerLYsAspGluAsn..... 1304
5472 GAAGGCTTACAAAGCCGAGCTGAAGAGCGGAGAAAGAGCCACACCA 5521
1305 .....GlnHisSerIleIleGlyGlnArgGlnGlu 1313
5522 ACCTGTCCAAAGTCCGCAAGCTGCACAGAGCTGATGAGCAGAGAG 5571
1314 IlePheAsp 1316
5572 AGCGCGGAC 5580
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqn/NA1995.DAT:AA086851

seq_documentation_block:
ID AA086851 standard; DNA; 8789 BP.

AA086851;

27-FEB-1996 (first entry)

Human mitosis gene.

Cell cycle; M phase; mitosis; retinoblastoma; mitosis; cell growth;
inhibition; ss.

Homo sapiens.

MO9511309-A2.

27-APR-1995.

24-OCT-1994; 94WO-US12162.

22-OCT-1993; 93US-0141239.

(TEXA) UNIV TEXAS SYSTEM.

Lee W, Zhu X;

WPI; 1995-170229/22.

P-PSDB; AAR72826.

Purified mammalian protein mitosis and agents that bind it and
inhibit its action - used to promote cell growth or to inhibit cell
division and/or proliferation

Claim 8; Fig 8B; 61pp; English.

AA086851 is the human mitosis gene. Mitosis is involved in the
regulation of the mammalian mitotic cell cycle. Mitosis as with E2F-1
(see AAR72824) interacts with the retinoblastoma protein (the
retinoblastoma tumour suppressor gene product). Mitosis is first
synthesised at the G1/S boundary, it is then phosphorylated from S
through M phase, and during mitosis, is closely associated with the
centromeres/kinetochores at the mitotic spindle poles. Mitosis is
necessary for a eukaryotic cell to enter the M phase of the mitotic cell
cycle and its degradation is necessary for a cell to advance on to the
next stage. Mitosis is thus useful for controlling cell growth as
overexpression of mitosis prevents a cell from exiting the M phase.
An anti-mitosis antibody, antibody fragment or a phosphorylated mitosis
mutagen (or nucleic acid encoding it) can also be used to inhibit cell
division which is particularly useful for the study of the cell cycle.

CC A further use is to control hyperproliferative cells, and so control
CC diseases such as psoriasis and breast cancer. It can also be used to
CC block gametogenesis of an immature gamete.

XX Sequence 8789 BP; 3169 A; 1517 C; 2069 G; 2034 T; 0 other;

alignment_scores:

Quality: 439.50 Length: 1256
Ratio: 0.668 Gaps: 49
Percent Similarity: 52.389 Percent Identity: 19.904

alignment_block:

US-09-538-396-2 x AA086851 ..

Align seg 1/1 to: AA086851 from: 1 to: 8789

51 CysThrGlyGluLeuProPheAsnSerArgSerGlyHisThrPheValAla 67

3808 TGCTCTGGGGAGACGTCGCCAGATACCAAT.....TA 3839

67 sAspProLYsValAlaGlyGluThrGluThrLYsGlyGlnIleLYsLeuA 84

3840 TGAGCCT.....CCAGGGAGAGATAAACCCAGGCTCTTCAGAA.... 3879

84 rPheLYsThrAlaAlaGlyLYsAspValValCysIleArg....Ser 98

3880TGCAATTCCTGAATTCGA 3897

99 PheGlnLeuThrGlnLYsAlaSerLYsMetGluPhe.....LY 111

3898 TTTTCTGGCTCTAATGCTTGTGACCTATGAGATTTCTCGGGAGATCAGA 3947

111 sAlaIleGluSerValLeuGlnThrIleAsnProHisThrGlyGluLYsV 128

3948 AGATATCCAAATATCTCACTGCGGGTAAAGAGACATCAAAATGAGAAT 3997

128 aLYsLeuSerTYrArgCysAlaAspMetAspArgGluIleProAlaLeu 144

3998 TGAGATTACTTCACTGATAGAGAGCCGTCACGAAAGTGAACGTTTG 4047

145 MetGlyVal.....SerLYsAlaValLeuGluAsnValAla 156

4048 CTAAATGAATGAAGAATTAGACTCAAAATCTCATTTACAGAGATACA 4097

156 ePheValHisGlnAspLYsSerAsnTYrProLeuGlnAsp...ProSerT 172

4098 ACTAATGACCAAAATGGAAGCATGATAGAAATGGAAAAATAGTTGGGG 4147

172 hrLeuLYsLYsLYsPheAspAspIlePheSerAlaThrArgTYrThrLYs 188

4148 AACCTTAAGAAAGAACTCAGATTTA.....ACTGAA 4179

189 AlaLeuGlnValIleLYsLYsLeuHisLYsAspGlnMetGlnGluIleLYs 205

4180 AAATTGGAATATTTCTTGTGATCCAGAGAGATTCCTCCAGAGATAGA 4229

205 sThrPheArgLeuLYsLeuGlnAsnLeuGlnThrValLYsAspGlnAlaH 222

4230 AACCTTGGAAGGCTCATCTGATTTAGAAATGCATGCGAATAAATCAT 4279

222 IsLYsLeuArgGluAsnIleAlaGlnAspGlnLYsSerAspAlaSer 238

4280 CA.....CCTGAAGATATGAGAAATGCGCAAGGTGAATGACACG 4323

239LYsSerGlnMetGlnGlnIleuLYsG 247

4324 TGAGAGAGAGATTTCTGTGATGTGAAATGACCTGATGATCGATCAGATC 4373

247 ulYsIleCysGlyThrGluArgGluIleLeuGlnMetGluThrSerLeuA 264

4374 GGAGAAAGCATGATGAGCATGAGAGCCCTACGCGAGAGCGTGAAGCTTGG 4423

264 spgIleuArgArg..... 268
4424 AGTAGTTCACACAGACAGATGTTAGAAAAAGCAATGAAATAG 4473
269LeuGInGlyGlnIleAspIleuValThrG 279
4474 CAGAGGTTATTGTCGCTGGAAGAAAGACTCAGTGCACAGAGTA 4523
279 uArgSerThrIleu.....LeuThrGlnGlnIleGlu 290
4524 GAGAAACCGCTGCTGGAGATTAGATCTATGTCAAAAAACCCAGG 4573
290 yIleuAlaIleuSerGlnGluAsnGluAspThrAspGluIleuMet 306
4574 CACTGATCACTTGTCTGAAAAATGAAAGGAAAAACACAGACTTCAG 4623
307 GlnTPrGlnThrLysPheGluIleuValIleuLeuGlnThrLys 323
4624 TCTCATCAAAAGTGAAGTGTCTCCATTCATTCAGGTGCAGCAGAGGT 4673
323 eSerLysLeuValIleuAspMetAspAspGluAlaSerLysSerVal 340
4674 GAAGGAAAAAGCGAACTCTTCAGACTTGTCTCTGATGTGAGTGAGC 4723
340 euserLysGlnAsnSerGluLeuThrHisGlnIleGlyLysLeuGlnAla 356
4724 TGTAAAAAGCAAACTCATCTCCAGGAAAGCTGCAGAGTTTGAAAG 4773
357 GlnAlaAspAlaHisLeuThrMetLysHisGlnIleuValAspSerAspIleu 373
4774 GACTCAGCAGGACGTCTTTCACAAATGTGAGCTGGAAGCAAAATGTC 4823
373 sasnIleCysThrLysHisAsnLeuGlyProValProGlnHisProPheT 390
4824 ACAACTGATTAAGAG..... 4839
390 hAsnAspValAlaMetAsnLeuThrAsnArgIleLysAlaArgLeuSer 406
4840 ..AAAGATTGCTTCTCAAGAACTGAAACCTGCAGCCAGCTGACT 4887
407 SerLeuGlnAsnAspLeuAspLysLysSerAsnGlnAspGlnLe 423
4888 GAATCAGATTGTGAAGAACTGAAATGCTCCAAAGCCTTGAGCGCGACT 4937
423 uAspValLeuThrLysHisLysLysLysLysLysLysLysLysLysLys 440
4938 GGTGAGAAAGGTGAGTTCGATTCAGCTGAGCTGACACAGAGGAG 4987
440 aAspGlyGlnIleGlnSerLysIleGlnSerMetSerGlyIleuValArg 456
4988 TGCAT...CAGCTGAGAGAGGATCGAAACTGAGA.....GTTCGC 5028
457 ArgArgLysAspLysGlnLysGlnIleuValAspAlaIleuValIleu 473
5029 ATTGAGGCCGATGAAGAAAGAGAGCTGCACATCGCAGAG..... 5067
473 rLysPheAsnLeuSerArgIleAspGlnArgIleuValHisMetGlnIleG 490
5068AAACTGAAGAACGCGAGCGG..... 5088
490 LuValGlnArgLysThrIleuAlaLeuGlnGluArgAspLysAspSerIle 506
5089GAGAAATGAT..... 5097
507 IleSerGlnLysArgThrGluValLysSerLeuGlnGlnLysLysVal 523
5098TCACTTAAGATAAAGTGAGAA 5120
523 IleuLeuArgGlnLysAspIleIleAsnArgAsnAlaAspGlnVal 540
5121 CTTTGAAGGGCAATGTCAAGATGTCAAGAAAGAAC...CAGGAGCTAGTGA 5167
540 yLeuGlnLysLysAspAlaLeuGlnUserSerLysAspLysLeuAsn 556
5168 TTCTT.....GATGCC...GAGATTCCAAAGCAGAGTAGAG 5202
557 GlnIleValAsnGlnHisLysAspLysIleLysLysValIleuArgLys 573
5203 ACTTA.....AAACACAAATGAGAGAGTGGCCAGAGGCT 5240
573 gAsnProPheGlnLysAspMetLysLysGlnIleAsnGlnAlaPheTrp 590
5241 GAAGTTTTCATTAAGCTTCTGC..... 5265
590 roValAspLysGlnLysAsnGlnLeuArgSerLysSerGlnAlaGlu 606
5265ACGTTAAGCTCTGAAAAAGAAACTGACA 5295
607 GlnIleuLysPheThrGlnSerLysValThrAspAlaArgGlnIle 623
5296 AAACAAATACAAAGAAACAAAGT.....CAGTT 5324
623 uThrLysLeuArgArgAspMetAspAlaLysArgArgPheLeuAspSerL 640
5325 CTCAGAACTACAAAGTTACTCTCTCTTAAAGTCTGTTAAGAGAA 5374
640 yLeuGlnSerIleLeuGlnIleSerAlaAsnValAspMetPheProLys 656
5375 AGAGCAAGCAGAGATACAGATC..... 5397
657 ValLeuGlnAspAlaMetAsnLysArgAspGlnGlnLysAlaGluAs 673
5398AAAGAAATCTAAACCTGACGTGAGAT 5426
673 nPheAlaAsnGlnLysMetArgLysMetLeuAlaProPheGlnHisLeuAla 690
5427 GCTTCAGAAATCAGTTAAAGAGCTAAATGAGCAGTACAGACCTTGTCTG 5476
690 rGlyAsnHisValCysProCysGlyArgAlaPheThrProAspGlu 706
5477 GTGACCAAGAAATTAAGAAAGCCACAGAAAGAGCTGAGCCACCAATA 5526
707 GlnAspGlnPheValLysLysGlnArgMetGlnAsnSerSerThrAlaG 723
5527 GAGGAAGG.....CATCAGCTGAGAAATACATTAAGAAAGCT 5564
723 uArgSerLysAlaLeuAlaMetGlnUserSerAsnAlaGlnAlaLeuPheG 740
5565 GAGAGCCCGCTCAGAGCTGATGAAGAAAGACAGCTC...TGTGCTTAC 5611
740 InGlnLeuAspLysLeuArg.....ThrIleuAspAlaIleuVal 753
5612 AACAACTGAGAAAGTAGCATCATGACAGATTACTTAAGGCTAGAGTG 5661
754 LysLeuValIleuGlnThrIleProLeuAlaGlnLysAsnLeuAsnGlnH 770
5662 GAGAACTTGAAGAGAGCTGAGATACCGACCAAC...CAGAGACA 5708
770 sLeuAlaAspGlnUserGlnLysAlaGlnAlaPheAspLeuGlnLys 787
5709 TGCAGCTTTTGAAGCGAGAAATTCAAA..... 5736
787 AlIleuAlaHisValGlnMetAspArgAspAlaValIleuAlaLeuGln 803
5737GAGAGAGTAGAGACCTTAAGCAAAATGAAAGGATGACCA 5781
804 ProThrAspThrIleAspArgHisVal..... 812
5782 AGCTGAGAGCTGGAATGATGATGTTTACTTAAGCTCAGAAAAAGA 5831
813HisGlnIleGlnIleuValLysGlnValGlnAspLeuG 826
5832 AAATCTGCAATGATTAACAAAGAGACCAAGCGAATATCTGATTAAG 5881
826 LuThrAlaLeuAspSerSerLysArgGlyVal.....LysSerLeu 839


```
473 rlyspheasnleuserarglleaspgluargluarghismetglnleag 490
508 .....AACTGAAGAACGCCAGCGG..... 5088
490 luvalgluargllythrleualaleugluargaprtyraspserrile 506
5089 .....GAGATGAT..... 5097
507 lleserghlnlysarthrghlnvaltyrserleuglnlyllyllysva 523
5098 .....TCACCTAAGATTAAGTGAAGAA 5120
523 lleuuarpgluulysaspilleleasnargasnllaaspgluarval 540
5121 CCTGAAGGGAATTCAGATGTCAGAGAAAC...CAGAGCTAGTA 5167
540 yslenglyleuylsaspalaleugluserSerlysnrlysplyseuan 556
5168 TTCTT.....GATGCC..GAGAAATCCAAAGCAGACAGTAGAG 5202
557 gluilevalasnglnllysnrlyllyllysvaalleuarglyar 573
5203 ACTCA.....AAACACAAATAGAAAGATGCCAGAACCT 5240
573 gasnrproheglulysaspmetlylsglulileasnglnlalepnetrp 590
5241 GAAAGTTTTGAATTAGACCTGTC..... 5265
590 rovalasplysglulyrasnnglnleuarargserlyserglnlualaglu 606
5266 .....ACGTTAAGCTCGAAGAAAGAAATCTGACA 5295
607 glnlgluleuylspherhrglnserlyvalthrapsalalearglgnle 623
5296 AAACAAATACAGAAACAAAGT.....CAGTT 5324
623 uthrlyseuarargaspmetaspalaleysargargpheleuasperl 640
5335 CTCAGACTGACAAAGTACTCTTCATTAAAGTCTTGAAGAA 5374
640 yslenglnserlleuglnlleserlaleasvalaspmethheprrlys 656
5375 AGGACCAAGCAGATACAGATC..... 5397
657 valleuglnaspalametasnlylsarargspglunglnlysargleu 673
5398 .....AAGAAAGAACTAAACCTGAGTGGAGAT 5426
673 nphelalasnnglymetargglumetleualarproheglulhlsleual 690
5427 GCTTCAGATCACTTAAGGAGCTTAATGAGCAGTACGAGCCTGTGTG 5476
690 rgluasnhisvalcysprocyssyluaralapherhproaspglu 706
5477 GTGACCAGAAATTTATGAAGGCCACAGAACAGCTTAGACCCACCAAA 5526
707 gluaspgluarhevallylsglnarymetglnasnserserthrilacl 723
5527 GAGGAAGAG.....CATCAGCTGAGAAATAGCATTTGAAGAT 5564
723 uargerlysalaleualametgluserseasnalaaglualeuphg 740
5565 GAGACCCCGCTAGAGCTGATGAAGAAAGACAGCTC..TGTCCTTAC 5611
740 lnglnleuasplyseuar.....Thrleryasplaryval 753
5612 AACCACTGAGAGAAAGTGCATCATGAGATTACTTAAGGCTAGAGTG 5661
754 llyseuvalaglnluthrlleproeualagluylsasnleuasnglnh 770
5662 GAGAACCTTGAAGAGAGCTAGACATAGCCAGACAAAC...CAAGACA 5708
770 sleualaspgluserglnlysalaglnlaleaspsapleuengly 787

5709 TGCAGCTCTTAGCGCAGAGAAATCCAA..... 5736
787 alleualahisvalglmetasparargspalaleualaleuengln 803
5737 .....GGAAGCTAGAGACCTTAAAGCAAAATAGAAAGGATGACCAA 5781
804 prothrapsrthrleasparghisval..... 812
5782 ACTCTGAGAGCTGTGAAATTAAGATGTTTACTATTAAGCTCAAGAAAGA 5831
813 .....Hlsgluileglnlenglnleuvallylsgluvalgluaspleug 826
5832 AATCTGCAAAATGAAATTCAGAAAGAGCAAGAGCAAAATCTGAATTG 5881
826 luthralaleuaspserserlyargglyval.....lyserleu 839
5882 AA...ATAATAATCATCATTTGAAATAATTTGCAAGAAAGAGCA 5928
840 glugluileglnleuualasphelenglnargythrargaprrhle 856
5929 GAGAAAGTACAGATG.....AAGAAATTCAGAGC 5960
856 ullevalgluvalaspspleuarargspglnhlsarpmetleuasnglu 873
5961 TGCATGAGATG.....CTTCAACACAAATTAAGAGCTCAATGACA 6004
873 spmetsereralaglnvalargtrphlsasnalaargglulu...lys 888
6005 GAGTGCACCCCTG.....CATTAAGCAGAAAGACCGTAGAG 6042
889 vallysalaserSerlleuenglnarphleglnlyserglnlgnle 905
6043 GCCAAAGCAGAAAT..... 6057
905 ualleuualaglnlglulysglnlenglnleuvalglulyslyleu 922
6058 .....CTTAGAGTCAGATGAGATGCTTGAACCTGAGAGCTCAGT 6100
922 euenglnluserleuaspro.....leuserlys 931
6101 TGCACAAAGCCTTGATGAGGCCAAATAATTAATGTTTTCAGATCT 6150
932 gluylsglnserlleuenglnluthrasnalaaleuylsglnlyseus 948
6151 TCAGTGAATAGCCCTCATTCAGAGATGAGATGCGACAAACCTGGA 6200
948 pglugluthrasnleuualaglnludrlysarpglupheglnlgnl 965
6201 GAAAGAGATGAAAGATCAGTAGACTGAATAATTCAGAGCTGAG 6250
965 euaspalaleuagllyargleuasmetlyllylsglytyrleuasps 981
6251 AGCAGCTTGTG..... 6261
982 lyslyasnnglnlyleuylsglnlenglnlyarghisvalleucys 998
6262 .....TCTAACTGTCCAGGTGGAAGAGACCACTTGGAA 6302
998 sserglnleuglnsercymetalalynglnlenglnlleserlaleu 1015
6303 GAGGCAA.....AACTGAAGCTGAGAAATTCAGACAGTGAAT 6340
1015 euasnluserlysglnleuenglnlenglnlylenglnleuysarasn 1031
6341 TGGAGCAGAAAGATCCAGTGTCAATCCAA.....AATGCTCT 6381
1032 llespaspasnleuyls.....Tyraglylthrlysal 1043
6382 TTGACAGACACATTAAGAGTGTGACAGATCTTACAAAGCTGAGAA 6431
1043 asapvalglunglnleuthrargaprrllegluserleuglnlualargleu 1060
```


6432	TCAGCTTGAAATGCACAAAAAATGACAAAAATGCTCTTGTGAAAAA....	6477
1060	euserTlleglSerleuserAlalleglAlaAspLeuLysArgHisSer	1076
6478GTAAACAAATGACTGCMAAGAAAGAACTGACGAGCGGAAATG	6522
1077	GINGLU.....LysGIuArgLeuAsnSerGI	1085
6533	CATGAGATGGCACAGAAAAACAGACAGCTGCMAAGAAAGAACTAGTGGAGA	6572
1085	uPheAsnArgTrpGlnGlyThrLeuSerValTyrGlnSerAsnIleSerL	1102
6573	CAAAAATAGCGTACGCTGGAGAGATTGCAGTACTGTTGGAGAAATTAAGA	6622
1102	YsnHLeysGlnGluLeuLysIleUserGlnTyrLysAspIleGluLysArg	1118
6623	GCAGCAAAAGATCAATTGAA.....	6642
1119	TyrThrAsnGlnPheLeuGlnLeuLysThrThrGluMetAlaAsnLysAs	1135
6643GACCTCACACTAAGAAATAGTGAA.....	6666
1135	PheAsnArgTyrTyrThrAlaLeuAspLysAlaLeuMetAlaAsnLys	1152
6667TTGAAGAGAGCCTAGATTGCATGCACAA	6694
1152	erMetLysMetGlnGluIleAsnLysIleIleLysGlnLeuTrpGlnGln	1168
6695	AAGACACAGGTGGAAAAAGGAAGTAAGTAGAGAGAAATAGCTGAAATAT	6744
1165	ThrTyrArgGlyGlnAspIleAsp.....TyrIleSerIleAsnSerAs	1183
6745	CAGTACGCGCTTCATGAAAGCTGAAAAAGAAACACAGGCTTCTTTTGA	6794
1183	PserGlnGlyAlaGlyThrArgSerTyrSerTyrArgValValMetGlnT	1200
6795	CACA.....AACAAACAGTATGAGTAGAAGAAATCCACA	6826
1200	hGlyAspAlaGluLeuGluMetArgGlyArgCysSerAlaGlyGlnLys	1216
6827	CATACCGAGAGAAATTAGCTCTTAAGAAGAAATGTCTCAGTTCACAGAG	6876
1217	ValLeuAlaSerLeuIle 1222	
6877	CTGGAGATGAGACCTTTTA 6894	

seq_name: /SIDSL/gcdata/geneseq/geneseq/NM196.DAT: AAT34578

seq_documentation_block:

ID: AAT34578 standard; CDNA to mRNA; 10136 BP.

XX AAT34578;

XX AC

XX 08-OCT-1996 (first entry)

XX DE Kinetochore protein CENP-F CDNA.

XX KW Kinetochore protein; CENP-F; cell cycle; cancer; diagnosis; ds

XX OS Homo sapiens.

XX Key Location/Qualifiers

XX CDS 171..9917

XX FT /*tag= a

XX WO6617867-A1.

XX 13-JUN-1996.

XX 08-DEC-1995; 95WO-US16216.

XX 09-DEC-1994; 94US-0353700.

PA (FOXC-) FOX CHASE CANCER CENT
XX (UYTE-) UNIV TECHNOLOGIES INT INC.
PT
XX Rattner JB, Yen TJ;
XX
DR PPI; 1996-287116/29.
DR P-PDB; AAR9795.
XX

DNA encoding kinetochore protein - used as a marker for the G2 and M phases of a cell cycle, partic. for detection of malignant diseases

Claim 6; Page 55-62; 72pp; English.

CC A cDNA sequence (AAT34578) encodes a 372 kDa human kinetochore protein, CENP-F (AAR9795), that is detected by immunofluorescence microscopy only during the G2 and M phases of a cell cycle. The sequence was deduced from clones isolated. from a human breast carcinoma cDNA library in lambda gill by screening with serum contg. CC CENP-F antibodies and by 5'RACE. The cDNA can be used for prodn. of CC large quantities of recombinant CENP-F, or to design probes for the CC detection of CENP-F genes in test samples as a means of monitoring cellular proliferation.

SX
XX Sequence 10136 BP; 3767 A; 1768 C; 2396 G; 2205 T; 0 other;

Alignment_scores:
Quality: 439.50 Length: 1256
Ratio: 0.668 Gaps: 49
Percent Similarity: 52.389 Percent Identity: 19.904

alignment_block:
US-09-338-396-2 x AAT34578 ..

Align_seg 1/1 to: AAT34578 from: 1 to: 10136

51 CysThrGlyIleuProPheAsnSerArgSerGlyHisThrPheValHl 67
||||| | | | | | | | | | :
5619 TGCTGTGGGGAACAGTCCCGCATGCCAAT.....TA 5650

67 sAspProLysValAlaGlyIuThrGluThrLysGlyGlnIleLeuA 84
||| | | | | | | | | :
5651 TGAGGCT....CCAGGGAAAGATAAACCCAGGGCTTCAGAA... 5690

84 rGpHelysThrAlaAlaGlyLysAspValValcysIleArg....Ser 98
||| | | | | | | | | :
5691TGCAATTTCGTAATGTCA 5708

99 pHeGInLeuthrGInLysAlaSerLysMetGluPhe.....Ly 111
||| | | | | | | | | :
5709 TTTTGTGGCTCTAATGCTTTGGTACTATGAGATTCTCGGGAAATCAGA 5758

111 salAlleGluSerValLeuGlnThrIleAsnProHisThrGlyGlnLysV 128
||| | | | | | | | | :
5759 AGATATTCATATCTTCACTCGGGCGGTAAAAAGACATCAATATGAGATT 5808

128 aLcylSeuSerTyrrArgCysAlaAspMetAspArgGlnIleProAlau 144
||| | | | | | | | | :
5809 TAGATTTACTTCATGTGATGAGAGACCGTAGCAAAAAGTTGAAGTTTG 5858

145 MetGlyVal.....SerLysAlaValLeuGlnAspValIl 156
||| | | | | | | | | :
5859 CTAAATGGAATGAAAGAAATAGACTCAAAAATCCATTATTCAGAGAGTCA 5908

156 pHeValHisGlnAspGluSerAsnTrpProLeuGlnAsp...ProserT 172
||| | | | | | | | | :
5908 ACTAATGCACAAATGTAAGCATGCATAGATTGGAAAAAATAGTTGGGG 5958

172 hLeuALyLysLysPheAspApIlePheSerAlaThrArGTrThrLys 188
||| | | | | | | | | :
5959 AACCTAAGAAAGAAACACAGATT.....AGTGA 5990

189 AlaleuGluValIleLysLysLeuHisLysAspGlnMetGlnGlnIleLy 205

```
||||| : : : : : ||| : : : : : ||| : : : : :
5991 AATTGGATATTTCTTGATCATCCAGAGTTACTCCAGAGTAGA 6040
205 SThrPheArgLeuLysLeuGluAsnLeuGlnThrValLysAspGlnAlaH 222
: : : : : : : : : : : : : : : : : : : : : : : : : : :
6041 AACTTCTGAGGCCCTCAATTCGATTAGAAATGCATGCAGATAAATCAT 6090
222 LsLysLeuArgGluAsnIleAlaGlnAspGlnGluLysSerAspAlaSer 238
||||| : : : : : ||| : : : : : ||| : : : : :
6091 CA.....CGTGAAGATATTGGAGATAAATGTGGCCAGAGTGAATGACACC 6134
239 .....LysSerGlnMetGluGlnLeuLysGln 247
6135 TCGAAGAGAGAGATTCTTGATGTGAAATGAGCTGAGTAGATCAGATC 6184
247 uLysIleCysGlyThrGluArgGluIleLeuGlnMetGluThrSerLeuA 264
: : : : : ||| : : : : : ||| : : : : : ||| : : : : :
6185 CGAGAAACCTAGCATTCAGATGAAAGCCCTTACCTGAGGCTGACTAG 6234
264 SPGLuLeuArgArg..... 268
6235 AGTAGATTCAACAGAGAGCTATTGTAGAAAAGACAAATGAAATAAG 6284
269 .....LeuGlnGlyGlnIleAspIleLysAlaThrGln 279
6285 CAGAGGTTATTGTCTGCTGCTTGAAGAAGACTCTAGTGTGACAGAGTGA 6334
279 uArgSerThrLeu.....LeuThrGlnGlnHisGluL 290
||||| : : : : : ||| : : : : : ||| : : : : :
6335 GAGAAACACGCTTCGTGAGAAATGATGACTATGCAAAAAAACACAGC 6384
290 yLysLeuAlaIleuSerGluGluAsnGluAspThrAspGluGluMet 306
||||| : : : : : ||| : : : : : ||| : : : : :
6385 CACTGTGATCAGTGTCTGAAAAAATGAAGAGAAAAACACAGAGCTGTAG 6434
307 GluTrpGlnThrLysPheGluGluArgGluIleAlaLeuLeuGluThrLys 323
||||| : : : : : ||| : : : : : ||| : : : : :
6435 TCTCATCAAGAGAGTGTCTCATTCATTCAGGTGCGCAGAGCAGAGCT 6484
323 eSerLysLeuValArgAspMetLysAspGluAlaSerLysSerValL 340
||||| : : : : : ||| : : : : : ||| : : : : :
6485 GAAGAAAAAGAGCGAATCCTTCAGACTTGTCTGTGATGTGAGAGAC 6534
340 euSerLysGlnAsnSerGluLeuThrHisGluIleGlyLysLeuGlnAla 356
||||| : : : : : ||| : : : : : ||| : : : : :
6535 TGTAAAAAGACAAAACATCATCTCCAGGAAACCTCAGAGTTGAAAAAG 6584
357 GluAlaAspAlaHisLeuThrMetLysHisGluArgAspSerAspIle 373
6585 GACTCAGAGGCACTGCTTTGACAAAATGTGAGCTGGAACCAAAATTGC 6634
373 sAsnIleCysThrLysHisAsnLeuGlyProValProGluHisProPhe 390
||||| : : : : : ||| : : : : : ||| : : : : :
6635 ACAACTGAATAAGAG..... 6650
390 hAsnAspValAlaMetAsnLeuThrAsnArgIleLysAlaArgLeuSer 406
||||| : : : : : ||| : : : : : ||| : : : : :
6651 ..AAAGATTGCTCTCAAGGAATCTGAAGCCCTCAGGCCAGAGTGA 6698
407 SerLeuGluAsnAspLeuLeuAspLysLysSerAsnGluAspGlnLe 423
6699 GAATCAGATTATGAAAAGCTGAATGCTCCAAAGGCCCTTGAGGCCGCACT 6748
423 uAspValLeuThrLysHisLysLysLysLysLysLysLysLysLysLys 440
6749 GGTGAGAAAGAGTACGTCGATTCAGCTGAGCTGACACACAGAGAGAG 6798
440 aLAspGlyGlnIleGlnSerLysIleGluSerMetSerGlyLysLeuArg 456
6799 TGCAT...CAGCTGAGAGAGAGCATCGAAGAACTGAGA.....GTTCGC 6839
457 ArgArgLysAspLysGluLysGluArgAspAlaIleGluValGluLeu 473
: : : : : ||| : : : : : ||| : : : : :

6840 ATGAGCCGATGAAAAAGACAGCTGCACATCGCAGAG..... 6878
473 rLysPheAsnLeuSerArgIleAspGluArgGluArgHisMetGlnIleG 490
6879 .....AACTGAAAAGACGCGCGG..... 6899
490 LuValGluArgLysThrLeuAlaLeuGlyGluArgAspTyrAspSerIle 506
6900 .....GAGATAT..... 6908
507 IleSerGlnLysArgThrGluValTyrSerLeuGlnGluLysIleLysVal 523
6909 .....TCACCTTAAGATAAAGTTGAGAA 6931
523 lleuLeuArgGluLysAspIleIleAsnArgAsnAlaAspGluValL 540
6932 CCTTGAAGAGAAATTCAGATGTCAGAGAAAGAAC...CAGGAGCTTAGTA 6978
540 yLeuGlyLysLysAspAlaLeuGluSerSerLysAspLysLeuAsn 556
6979 TTCTT.....GATGCC...GAGAAATTCAAAGCAGAAATGAGAG 7013
557 GluIleValAsnGluHisLysAspLysIleLysLysValLeuArgGlyArg 573
7014 ACTCTA.....AAACACAAATGAAAGAGATGCCAGAAAGCT 7051
573 gAsnProPheGluLysAspMetLysLysGluIleAsnGlnAlaPheTyr 590
7052 GAAGATTTTGAAATGTAGACCTGTCT..... 7076
590 roValAspLysGluTyrAsnGluLeuArgSerLysSerGlnGluAlaGlu 606
7077 .....ACGTTAAGGCTGAAAAAATAATCTGACA 7106
607 GlnGluLeuLysPheThrGlnSerLysValThrAspAlaArgGluGlnLe 623
7107 AAACAAATCAAGAAAAACAAGT.....CAGTT 7135
623 uThrLysLeuArgArgAspMetLysAlaLysArgArgPheLeuAspSerL 640
7136 GTCAGAACTAGAACAAAGTACTCTTCATTTAAAGTGTGTAAGAAAG 7185
640 yLeuGlnSerIleLeuGlnIleSerAlaAsnValAspMetPheProLys 656
7186 AGGAGCAGACAGATACAGATC..... 7208
657 ValLeuGlnAspAlaMetAsnLysArgAspGluGlnLysArgLeuGlnLys 673
7209 .....AAAGAGATCTAAAACTGCAGTGAGAT 7237
673 nPheAlaAsnGlyMetArgGluMetLeuAlaProPheGluHisLeuAla 690
7238 GCTTCAGATCAGTTAAAGAGCTAAATGAGCAGTAGCAGCTGTGTG 7287
690 rGlyLysHisValCysProCysGlyArgAlaPheThrProAspGlu 706
7288 GTGACCAGAAATTAATGAGGCCAGAGACAGCTGAGCCACCAATA 7337
707 GluAspArgPheValLysLysGluArgMetGlnAsnSerSerThrAlaGln 723
7338 GAGGAAGAG.....CATCAGCTGAGAAATAGCATTTGAAAGAGCT 7375
723 uArgSerLysAlaLeuAlaMetGluSerSerAsnAlaGluAlaLeuPheG 740
7376 GAGAGCCGCCCTAAGAGATGATGAAAGAGAGAGCTC...TGTGCTTAC 7422
740 InGlnLeuAspLysLeuArg.....ThrIleThrAspAlaTyrVal 753
7423 AACCACTGAGAGAAAGTGCATCATGAGATTTACTTAAGGTGAGAGTG 7472
754 LysLeuValGluGluThrLeuProLeuAlaGluLysAsnLeuAsnGlnI 770
7473 GAGAACTTGAAAGAGAGCTAGAGATAGCCAGAGCAAAAC...CAAGAGCA 7519
```

[illegible]

```

1043 aaspyaVgIuGluInleuthrArgspRIegIuseRleuGluGluValArgLeuL 1060
8243 TGACCTTGAATTGACAAAAATGGCAAAATGTCCTTTGTTGAAAA... 8288
1060 euseRIegIyseRleuseRAlaIlegIuaIaspleuysArghIisSer 1076
8289 .....GTTAACAAATGACTGTCGAAAGGAAMCTGAGTCGACAGGGAATG 8333
1077 GInGuL..... LysGluArgLeuAsnSerL 1085
8334 CATGAGATGGCACAGAAAAACAGACAGACTTCAGAAAGAACTCAGTCAGCA 8383
1085 uPheasnaArgTrpGlnIlythrIleuseRValTYrGInserAsnIleSerL 1102
8384 GAAAAATAGGCTAGCTGAGAGTTGCGTTTACGTTTGGAAGAAATTAAGA 8433
1102 ySHIstLySGInGluLeuLyseuseRGIInTYrLysaspIleGluLysArg 1118
8434 GCAGCAAAAGATCATTTGAG..... 8453
1119 TYrThrAsnGlnPheLeuGInLeuLYsthrThrGluMetAlaAsnLysAs 1135
8454 .....GAGCTCACACGTGAAATAGTGA..... 8477
1135 pleuAspArgTYrTYrThrAlaLeuAspLysAlaLeuMetArgPheHis 1152
8478 .....TTGAAGAGAGCCTTGATTCGATTCGATTCGACA 8505
1152 erMeTyseMeTGlUGluIleAsnLysIleIleLySGIuLeuTrpGInGln 1168
8506 AAGACCAAGGTGAAAAGAGGAAGAAATGACAGACAGAAATAGCTGAAT 8555
1169 ThrTYrArgGlyGlnAspIleAsp.....TYrIleSerIleAsnSerAs 1183
8556 CAGCTACGCGTTTCATGAAGCTGAAAGAAACACAGAGCTTCGTTTGA 8605
1183 pSerIuGlylaagIlythrArgSerTYrSerTYrArgValValMetGluT 1200
8606 CACA.....AACAAACAGTATGAAGTGAATCCACA 8637
1200 hrGlyAspAlaGluLeuGluMetArgIlyArgCysSerAlaGlyGlnLys 1216
8638 CATACCGGAGAAATGACTCTTAAGAAAGAAATGTCAGTTACACGAAG 8687
1217 ValLeuAlaSerLeuIle 1222
8688 CTGAGATAGACCTTTTA 8705

seq_name: /SIDSL/gcgdata/geneseq/geneseqn/NA1997.DAT:AA158840
seq_documentation_block:
ID ID AA158840 standard; DNA: 580073 BP.
AC AA158840;
XX
XX 27-MAR-1997 (first entry)
XX
XX Mycoplasma genitalium genome.
XX
XX M. genitalium; DNAA; DNA gyrase; origin of replication;
XX megabase shotgun sequencing method; open reading frame; ORF. ss.
XX
XX Mycoplasma genitalium.
XX
XX Key Location/Qualifiers
XX CDS 8552..9184
XX /tag=
XX /label= M006
XX /note= "Previously identified as MORF-20076, then
XX encoded protein shows 27.59 percentage
XX identity to thymidylate kinase (CDC8)
XX from Saccharomyces cerevisiae"
XX
XX CDS 11252..12040

```


495 ThrLeuAlaLeuGlyGluArgAspIleSerIleLeuSerGlnLysAr 511
260259 CACTTAGAACACACCAAAATGAGTTT 260285
511 GThrGluValIleYrSerLeuGlnGlnLysIleLysValLeuLeuArgGluL 528
260286 GAACAAA 260294
528 YAspIleLeuAsnArgAsnAlaAspGluArgValLysLeuGlyLeuLys 544
260295 CAGCTGTAGAGCTCTTAAGAGCTTGAAGCTGA 260327
545 LysAspAlaLeuGlnSerSerLysAspLysLeuAsnGluIleValAsnGlu 561
260328 TATTAAGGCTTACAA CACAAGATTAATGAATTTAAAAATGA 260368
561 uHisLysAspLysIleLysLysValLeu... ArgGlyArgAsnProPhg 577
260369 AAGTCCACTAAGACGAGGAACCTTTTAACCAAGAACGGGAATATTG 260418
577 LuLysAspMetLysLysGluIleAsnGlnAlaPheTrpProValAspLys 593
260419 AAAAA CGCAGGAATTTGACAGCTTTTAACCAAGACATCTTG 260462
594 GluTyArgnGluLeuArgSerLysSerGlnGluAlaGlnGlnGluLeu 610
260463 GAATATGACACCAACGT GAGTCAGTCACTACTACTCA 260500
610 sPheThrGlnSerLysValThrAspAlaArgGlnGlnLeuThrLysLeu 627
260501 AATATAAGCAGATGAAGTAAACACACTCCAAAACCTTAGATAGCTA 260550
627 rArgAspMetAspAlaLysArgArgPheLeuAspSerLys 640
260551 AAAAGAGCTAGACAAAGAAGAACTCTTAGTACACAAAAAAAGTT 260600
641 ...LeuGlnSerIleLeuGlnIleSerAlaAsnValAspMetPheProL 656
260601 GATAGTGAAGCAATCTTCACTCAAGAAAGTGTCAAGAACAGTAA 260650
656 sValLeuGlnAspAla... MetAsnLysArgAspGlnGlnLysArgLeu 672
260651 GGAGCTTGAAGACTATACCTGTGTTAAAAACAAAGAACAGTCAAGG 260700
672 LuAsn PheAlaAsnGlyMetArgGlnMetLeuAlaPro 684
260701 AAAATGAAGTGTGTTTGTGTAAGACAGTTAAACACACCAAGCAGAT 260750
685 PheGluHis LeuAlaArgLysAsnH 693
260751 TTTGAAGAAAGATTTGAGACTAACAACAGAGTTGTTGAACTAACA 260800
693 sValCysProCysGlnArgAlaPhe ThrProAspGlu 707
260801 T GCATTTGAAAGCTTCTTTATCAAACTTGAAGATTAAGAAA 260841
707 LuAspGluPheValLysLysGlnArgMetGlnAsnSerSerThr 721
260842 AGATTTTAACACTAAGCACAACAGATGGCAATGATTTTCCCACTA 260891
722 ...AlaGluArgSerLysAlaLeuAlaMetGlnSerSerAsnAlaGluAl 737
260892 AAAACGTGAAGTCAAG AGTGCCTGATTTTGAACT 260926
737 AleuPheGlnGlnLeuAspLysLeuArgThrIleTyArgAlaIleTyVal 754
260927 AATGTGCAAAAT 260939
754 ystLeuValGluGluThrIleProLeuAlaGluLysAsnLeuAsnGlnHis 770
260940 GAGTATGAAAACTTGCAACAAAGAA 260963

771 LeuAlaAspGluSerGlnLysAlaGlnAlaPheAspLeuLeuGlyVa 787
260964 AAACAAAATTTTCCACAGACGTAATCTTGAAGAAAGCGCGCT 261013
787 IleuAlaHis...ValGlnMetAspArgAspAlaValGluAlaLeuG 803
261014 TTTATCAACCGGTTACACAAAACGT GAGGAGTTATTTAC 261054
803 InProThrAspThrIleAspArg 810
261055 AACAAAAGAAAACGCTGATCAGCTTACAAAAGCTTGAGCAAGACGG 261104
811 ...HisValHisGluIleGlnGlnLeuValLysGluValGluAspLe 825
261105 TTTATCAACCAAAAGGACACAAAGGATGTTGTCATCGTTGAAGAA 261154
825 uGluTyArgAlaLeu AspSerSerGlyA 834
261155 AAGAGATATTGGGCAAAAACCTCCAGATTTTCCCAACTCAGTGA 261204
834 rGlyValLysSerLeuGlnGluIleGlnLeuGlnLeuAsnPhe 848
261205 ACGCTCTAAAAATTAGCTGAACGGAGATGCAATCAAGTTTAAAGAA 261254
849 ...LeuGlnArgThrArgAspThrLeuIleValGluValAsp 861
261255 AAGGATATGAAGCACTGAAGAAACCACTGTTAATGATGTTAATATGC 261304
862 AspLeuArgAspGlnHisArgMetLeuAsnGluA 873
261305 TGAAGTTATCCAAAGCACTTTAGCAACACTCAACCAATCAATCAAG 261354
873 sPheSerSerAlaGln ValArgTrpHisAsn 883
261355 AACGCAGTAAATTGCAAAACGCCAAACAAAGATTGCTGATTTTCCAT 261404
884 AlaArgGlnGluLysValLysAlaSerSerIleLeuGlnArgPheGln 900
261405 GATTCCTAAAAAACTCAATGATGATGAACTT ACCTTCAAAA 261448
900 sSerGlnGlnGluLeu...ValLeuAlaGlnGlnLysGlnGlnLeu 916
261449 ACGGTGCAAGAAATTACAAACCTTGAGGCTAACCAAAAACAACTTCA 261498
916 LeuAlaGluLysLysLeuGlnGlnSerLeuAspProLeuSerLysGlu 932
261499 ATCAAAATCAAGCTTACTTGAAGTGAAGTGAATTAAGTAAACAGAA 261548
933 LysGlnSerLeuLeu GlnGluTyArgsnH 942
261549 AAACAAAGCTTTTGAACCTTACGTAAAGAAACAAACATAGAGGTTGATGC 261598
942 AleuLysGlnLysLeuAspGlnGluTyArgHisGlnLeuAlaGluArgLys 959
261599 TATTAAACAAAGGTTGATGATTAACATCAAGCTTTAAATATGCAACAG 261648
959 rGluPheGlnGlnGlnLeuAspAlaLeuGlyArg LeuAsnMet 973
261649 CAGAGCTAGATAGAAAACCCATGATGATTAATATGCTTTTAAACCAT 261698
974 LysIle LysGlyTyArgLeuAspSer LysLysAs 984
261699 GATCGCGATCAAAAAGATGTACAGCAACCACTGAAACCTTTAAAGAGC 261748
984 nGluLysLeuLysGlnLeuGlnGlnLysArgHisValLeuCysHisSerGln 1001
261749 CCAAAAATTAATGATTTAGAA...CGTAGTGCACAGCTTGAAGAAACAA 261795
1001 euGln SerCysMet 1005
261796 GTGAGTTGCTGAAGAAATGTTGCTGTTTAAAGCCCATGCTGTAATAA 261845
1006 AlaLysGlnGlnArgIleSerAlaGlnLeuAsnLysSerLysGlnLeu 1022

261846 ACTAGTCACACCCAAAAGATTATGACTGACCAAAAAACAGAAAGTGA 2618955
:::||||| ::::|
1022 uGlnGlyGlnGlyLeuLeuLysArgAsnIle.....AspAspAsnL 1036
||| ::::|
261896 GCAAACCCAAAAGAAAGAAAGAACTAAAGATTGCTTTAGTGATCTACAAA 261945
:::|||||
1036 euLysTyrArgLysThrLysAlaAspValGlnGlnLeuThrArgAspIle 1052
|||||
261946 AAGACTATACAGGTTTTTGAATCCCAAAAGACCAAGAAATTTAGACAAATT 2619955
:::|||||
1053 GluSerLeuGln.....GluArgLeuLeuSerIleGlySerLeuSe 1066
|||||
261996 GAAGCTAAGCAACGCGAACTTCACAAAGTTAGCTGAAAAAAATATACGGT 262045
:::|||||
1066 AlaIleGlnAlaAsp.....LeuLysArgHisSerGlnG 1078
:::|||||
262046 CAAACTAGAACTGTGATTAACAGGTTTCAACGCGTGCAAAACAAACAAAG 2620955
:::|||||
1078 LuLysGluArgLeuAsnSerGluPheAsnArgTyrPdnLysThrLeuSer 1094
:::|||||
262096 ACACAGTCAACAGCTCAGCTAGTAGAAGCGGAACGGAACACACAGTTAAC 262145
:::|||||
1095 Val.....TyrGlnSerAsnIleSerLysHisLysGlnI 1106
:::|||||
262146 CTGTGACAAACTGCTTTCAACCAAGCTAATGATCTTTTAAACACAGC 2621955
:::|||||
1106 uLeuLysLeuSerGln.....TyrLysAspIleGlnLysAl 1118
:::|||||
262196 TGAACACCTCCACCAAAAAGATCCAAAGCTTCCACATGAGTTGAAAAAGC 262245
:::|||||
1118 rGlyTyrThrAsnGlnPheLeuGlnLeuLysThrThiGlnMetAlaAsnLys 1134
|||||
262246 Gt.....AACCAATCTTACGCTTTAAAAAGGAAAAAGTTGTTGCACAAA 2622855
:::|||||
1135 AspLeuAsp..... 1137
:::|||||
262290 GAGCAAGATCAACAAACGCAAGATCAGAGATCACTGACGCTTTAAACA 262335
:::|||||
1138ArgTyrTyrThrAlaLeuAspLysAlaLeuMetArgPheHis 1152
:::|||||
262340 GTTTGAAAAAGAAATATACGATTTTATGAACTTAAGAAAAAGC..... 262385
:::|||||
1152 ezMetyLysMetGlnGlnIleAsnLysIleIleLysGlnLeuTyrPdnGln 1168
:::|||||
262383GAACCTGAAGAGCTGAAAGAAATCAGAGAAAGTTTAAGTCAAAAGC 262425
:::|||||
1169 ThrTyrArg.....GlyGlnAspIleAspTyr 1177
:::|||||
262428 AACGTTGAATTAGAGAAAAAGAAAAAGAAAACTGGCTATCTATATTCACATA 262475
:::|||||
1177 rIleSer.....IleAsnSerAspSerGlnLysAl 1187
:::|||||
262478 TTTAAATTAAGTTCAACACACACACCCAAATTTAACCGGATCAACACTTAACA 262525
:::|||||
1187 laGlyThrArgSerTyrSerTyr...ArgValValMetGlnThrGlyAsp 1202
:::|||||
262528 GTCAGATGACAGACGTTCTATTATGAAGCAAAAAACCTCCAAACGCTTTAGT 262575
:::|||||
1203 AlaGluLeuGlnMetArgGlyArgCysSerAlaGlyGlnLysValIleuAl 1219
:::|||||
262578 AATGAAGCTAATCTTAA.....AAAGC 262605
:::|||||
1219 AserIleuIleIleArgLeu...AlaLeuAlaGluThrPheCysLeuAsn 1235
|||||
262601 CTTTAAATTAAGCGCTTAAGAAAGCTTGCATCAATCAATAACTCCAAA 262655
:::|||||
1235 yGlyIleLeuAlaLeuAspGln...ProThrThrAsnLeuAspGlyPro 1250
|||||
262651 AAGAACGTTAGCAACCAAAAACTAGAGTTTGATTAAGCGCGATGAACAA 262705
:::|||||
1251 AenAlaGluSerLeuAlaAlaAlaLeuLeuArgIleMetGlnAlaArgly 1267
:::|||||

```

262701 CAGAAAAAGAGTTACAGCAGCACTACTTACACATAGAAAGCTCAAGTT 262750
1267 sGlycInGlnAsnPh.....GlnLeuIleValIle. 1277
      ::::::::::::::|
262751 TGAAGAGCAAACTTGTACATTGAAAAAACAAGCCCACTACTGTGTATT 262800
1278 .....ThrHisAspGluArgPheAla.....His 1285
262801 AAATCACTGTGAAAAAACTTACTGTGTGAAAAAAAAGCCACTTAACCAAAAG 262850
1286 LeuIleGlyGlnArgGlnLeuAlaGlnLysTyrTyrArgValSerLysAs 1302
      ::::::::::::::|
262851 CTAGTGTGAACATAAAACTTATCCCAAAACGTATCTT...GCTAAATAAGAA 262897
1302 pGlnAsnGlnHisSerIleIleGlnSerGlnGlu 1313
      ::::::::::::::|
262898 TAAAGCTGAATCTCCAGCAACAACCTCCACACAG 262931
seq_name: /SIDSL/gcgdata/geneseq/geneseq/NA1991.DAT:AAQ10378
seq_documentation_block:
ID      AAQ10378 standard; cDNA; 5858 BP.
XX
AC      AAQ10378:
XX
DT      12-APR-1991 (first entry)
XX
DE      Encodes human 160kD mediator of inflammation protein.
XX
KW      Mediator of Imflammation; cytokine; Hodgkin's lymphoma; MRP-160; ss.
XX
OS      Homo sapIens.
XX
FH      Key      Location/Qualifiers
FT      CDS      134..4417
              /*tag= a
              /product= MRP-160
              /note= "mediator of inflammation"
FT
FT
EP412050-A.
06-FEB-1991.
26-JUN-1990; 90EP-0810481.
05-JUL-1989; 89GB-0015414.
(CIBA ) CIBA GEIGY AG.
Odlink KG, Tarcey L, Bruggen J, Wiesendanger W, Cerletti N;
Sorg C, Dewolf-Peeters C, Delabie J;
WPl; 1991-038913/06.
P-PSDB; AARI0534.
160 kD human polypeptide mediator or precursor of inflammation -
polyclonal or monoclonal antibodies to polypeptide treat and
diagnose chronic inflammation and hodgkins lymphoma
Claim 14; Page 32; 47pp; English.
mRNA isolated from human U32 cells was used to synthesise ds cDNA
which was cloned in phage lambda gtl1. The library was screened for
the sequence encoding the molecule which reacts with Mab1 C5 (see EP
162812), by successive screening with Mab1 C5 and cDNA inserts from
the commercially available cDNA libraries ML1005B and HL1029B. One
of the positive clones contained a 6kb cDNA fragment which was found
to encode MRP-160. The sequence was cloned in pUCKO to give a hybrid
vector which was used to transform E.coli DH5-alpha cells. A plasmid
yielding the correct restriction fragments was designated pMRP160.
Sequence 5858 BP; 1962 A; 1171 C; 1374 G; 1351 T; 0 other:

```



```
alignment_scores:
  Quality: 436.50      Length: 1111
  Ratio: 0.751        Gaps: 48
  Percent Similarity: 52.295      Percent Identity: 20.612
```

a11gment_block:
US-09-538-396-2 x AAO10378

Align seg 1/1 to: AAQ10378 from: 1 to: 5858

[illegible]

1937 AACAAAGAGAACTCGATGATGAT... GCTGTATGAAAGGCCAAACTGGG 1983
432 gtle..... 433
1984 GACTGCATCGCATCCACACAGCAGCGANTGGAAGACTGAAGTATCTT 2033
434 AsnIlaargTysSerGIuValaApGly 442
2034 TCAGCAAAAGGCTTGGAAACAGAGAGCGAGAAATTTGCTGAACCTMAAACA 2083
443 GlnIle..... GlnSerTylIleGlnSerMetSe 452
452 rgIYIleUaTgArGArGlyAspYsGIuYsGIuValaTgAspAlaIaG 469
2129 TTGCAAGATTCAAACAAGACTCTGAGGGGCTGCCATGCTAAAG 2171
469 lUValGIuLeuSerTylsPheAsnLeuSerArGIleAspGIuArgIuArg 485
2172 AGATGGAACCTCTGGAGGGCTAAACATGAAGT... ATTAAAGAAAG 2218
486 HlsmEgHlllGllValGIuValGlyThrLeuAlaIleuGIuValaTgAs 502
2219 GAAACACAGTCTGGAAACCATCAGTGCAGAACTGGCAACAGCAAGAAACCA 2268
502 pTYrAspSerIleIleSerGIuYsArGIhGluValTyrSerLeuGuc 519
2269 GCAT..... CTGCTGAATTAAGAAAGACACGTTAAACAATTAACAGGAA 2312
519 lUlySIleYsValLeuLeuArgGIuYsAspIleIleAsnIhArgAsnIa 535
2313 CTGAATTAAGGTA... AAGGAGCTAGAGGTCTCCAGCCCAATAGC 2356
536 AspGIuArgValYsLeuGIuLeuYs... lYsAspAlaIleuGlnSerSe 551
2357 AATGACAAACCAAGGTTATATGATTAATTTACATACAGCTCAAGGCTAC 2406
551 rLYsAspLYsLeuAnGIuIle..... ValAnGIuH 562
2407 TGAAGAAACCTCTGTGATCTTGATGCACTTCGGAACCCAGTTCGGAAG 2456
562 lUlysAspLYsIleYsLYsValLeuArgGIuYsAsnProPheGIuYs 578
2457 GTAAATCGGAAAGAAAGAACTTAGACACAGCTTGAGGCAAGCTGAGAA 2506
579 AspMetLYsGIuIleAnGIuAlaPheTrpProValAspLYsGIuTY 595
2507 CAGATTAAACAT..... TTAGAGATGGAAA 2532
595 rAsnGIuLeuArSerLYsSerGIuAlaIleGlnGlnIleuLYsPheT 612
2533 GAATCGGAAGTAGACAGGCTAGTAGCAATACAGAGAGGCTCCAGGGGA 2582
612 hrgInserLYsValThAspAlaArgGIuGlnLeuThrLYsLeuArgYg 628
2583 GAGAGCTAAAGCTTACAACTTCAGGAAAAATTTGATGAATCTGCTCA 2633
629 AspMetAspAlaYsArGArPheLeuAspSerLYsLeuGlnSerIle 645
2633 GTGAAGAGACTTGGAAAAAGAACTTCAGATTTTAAAGAAAAGTTTGC 2682
645 uGlnIleSerAlaAnValAspMetPheProLYsValLeuGlnAspAlaM 662
2683 TGAAGCTTCAAGAGAGCACTCTCTCTCAGAGAAATATGACAGAAACAG 2733
662 etaAnLYs..... ArgAspGIuGln... lYsArgLeuGlnAsnPhe 674
2733 TAAATTAAGTTACACCAAAAGAGAGAACGTTTAAACATGCTGCTTCTGAC 2782
675 AlaAnGIuYsArGIuMetIleAlaProPheGIuIleAlaArgLY 691
2783 TTGGAGAAAGCTGAGAGAAACTTAGCAGATATATGGAGGCAAAATTTTGAGA 2833

691 sasnhiValCyserocyScygluArgAlaPheThrProAspGluGlu 708
 2833 GAA..... GATGAGGAGAG 2849
 708 ArgLupheValys...LysGlnArgMetGlnAsnSerSerThrAlaGlu 723
 2850 AGCAGCTGATTAAGCAGAAAGAAACTGGAAT.....GAC 2887
 724 ArgSerLysAlaLeuAlaMetGlnSerSerAsnAlaGluAlaLeuPheG 740
 2888 ATTCACAAAATATGAGAGATGTGAGAGATGACTCTTCAGCTGACAAA 2937
 740 nGlnLeuAspLysLeuArgThrIleTyrAspAlaTyrValLysLeuValG 757
 2938 GATGACGATGATTAATGCT...CTGAAAGAAAGAGATGTAGAAATAATAC 2964
 757 LngLutThrIleProLeuAlaGluLysAsn.....LeuAsnGlnHis 770
 2985 AGCTAAACTTACAAAGGCTATATGAAAATGCAAGTTTCTGCAAAAAGT 3034
 771 LeuAlaAspGlnSerGlnLysAlaGlnAlaPheAspAspLeuGlyVa 787
 3035 ATGAGGACATGACTGTCAAAGCTGA..... 3061
 787 LLeuAlaHisValGlnMetAspArgAspAlaValGlnAlaLeuGlnP 804
 3062CAGAGCCAGCAAGAGCAGCTAAA..... 3085
 804 rGlnPArgThrIleAspArgHisValHisGluIleGlnGlnLeuValLys 820
 3086AAGCATGAGAGAGAAAGAAAGAAATTTGAGAGG 3118
 821 GluValGlnAspLeuGluTyrAlaLeuAspSerSerGlyArgGlyValLys 837
 3119 AAATTGTGAGACTGGAAGAAATGGAACAGC..... 3154
 837 sSerLeuGlnGluIleGlnLeuGluLeuAsnPheLeuGlnArgThrArg 854
 3154 3154
 854 sPThrLeuIleValGluValAspAspLeuArgAspGlnHisArgMetLeu 870
 3155CACACCAAGTCT 3166
 871 AsnGlnAspMetSerSerAlaGlnValArgTyrPheHisAsnAlaArgGlu 887
 3167 CAGGAG.....CTGAAAGCCAGATGAGAGAGCCACTTCTGA 3204
 887 u...LysValLysAlaSerSerIleLeuGlnArgPheGlnLysSerGluG 903
 3205 GACAAAGAACCAACATGTAAGAAATCTTACAGAACTCCAGAAAGAGC... 3250
 903 LngLutLeuValLeuAlaGluLysGlnGlnLeuIleValGluLys 919
 3250 3250
 920 LysLeuLeu.....GluGlnSerLeuAspProLeuSerLysGluLysG 934
 3251 ...CTGCTGAGACAGAGAGCAAGCTGAAGGCGCAGCGGAGAGAACAG 3297
 934 uSerLeuLeuGlnGluTyrAsnAlaLeuLysGlnLysLeuAspGluGlu 950
 3298 TGGCTTGTGCGAGAGCTGAGAGAGCTGAGAAAGCAAGCCGCAAGCA 3347
 951TyrHisGlnLeuAla 955
 3348 AAGCTGTCAAACAGCGAGAGATGCCATGACATATGAGACAGATGACC 3397
 956 GluArgLysArgGluPheGlnGlnLysLeuAspAlaLeuGlyArgLeuAs 972
 3398 AAAGAGAAAGACAGACACTGTGCTCTTGGAGGAGACACCAACCAACAA 3447

972 nMetLysIleLysGlyTyrLeuAspSerLysLysAsnGlnLysLeuLysG 989
 3448 TGCAGAACTACAGATGATGTGACACACTTAAGAAACAACTGAAAA 3497
 989 LuLeuGlnGlyArgHisValLeuCysHisSerGlnLeuGlnSerCysMet 1005
 3498 ATGTGAA..... 3505
 1006 AlaLysGlnGlnArgIleSerAlaGluLeuAsnLysSerLysGluLeu 1022
 3506GAGCTGACAAATCAAAAGAACTCT 3531
 1022 uGlnGlnGlnGlnLeuLysArgAsnIleAspAspAspLeuLysTyrA 1039
 3532 GACTGTGAGATCA..... 3547
 1039 rGlySerThrLysAlaAspValGlnGlnLeuThrArgAspIleGlnSerLeu 1055
 3548AAATGAGAAATTTAGAAAGAAATAGAAACCTA 3583
 1056 GluGluArgLeuLeuSerIleGly...SerLeuSerAlaIleGluAlaAs 1071
 3584 AAGCAGCAGCAGCTCAGAAAGTCCAGCAGCTTTCAGCGTTGCAA.... 3628
 1071 PLeuLysArgHisSerGlnGlnLysGluArgLeuAsnSerGluPheAsnA 1088
 3629GAGAGAAAGTTAACTGCGAGAGAGCTGGGGA 3662
 1088 rGTPrglnGlyThrLeuSerValTyrGlnSerAsnIleSerLysHisLys 1104
 3663 GAACGAGGAGCAGAGTCACAAAGTCATCAA..... 3691
 1105 GlnGlnLeuLysLeuSerGlnTyrLysAspIleGluLysArgTyrThrAs 1121
 3692AAGCTGAGAGAAAGAAAGATCTGTG.....CTCAATTA 3723
 1121 nGlnPheLeuGlnLeuLysThrThrGluMetAla...AsnLysAspLeuA 1137
 3724 TCAATTGTGTAATGAAAGAAAGAAATCCAGTTCCTTAAGAGCGCAG 3773
 1137 sPArgTyrTyrThrAlaLeuAspLysAlaLeuMetArgPheHisSerMet 1153
 3774 ATGAAGAGAAAGCTCTCCCTGCAAGAAATCCATCACTATCACTAGTGCCTTA 3823
 1154 LysMetGlu.....GluIleAsnLysIleIleLysGluLeuTrpG 1167
 3824 CTCACAGAAAGAGATGCCAGAGCTGAGAAACTGAGAAATGAGGTC.... 3868
 1167 nGlnThrTyrArgGlnGlnAspIleAspTyrIleSerIleAsnSerAspS 1184
 3869 .ACAGTGTCTCAGGGAGAAAGCGCTCTGCCAAGTCTTGCAATCA... 3913
 1184 ergLysGlyAlaGlyThrArgSerTyrSerTyrArgValValMetGlnThr 1200
 3914GTGTTCAAGACT 3925
 1201 GlyAspAlaGlu.....LeuGlnMetArg 1208
 3926 CTAGAGTCTGATAGGTAGTGAAGCTCGAGCTCAAG 3958

32 levaldylproasgylvalaglyushrthrllelleglucysleuys 48
::: |||||
489 tggttggacccaaatggggccgggaagacgacacatcatgtaattgctttaa 538
::: |||||
49 leusercysrthrglygluleuproproanserargsergluylstrph 65
::: |||||
539 tatatttctactggagaaatttccctcctcggaaccaa...ggaaatacat 585
::: |||||
65 evalhlhaspprocluyvalalaglyluuthrgluuthrlysglygluilel 82
::: |||||
82 yslleuargphelylestrhlalaglyluyasapvalalcyisleargser 98
::: |||||
636 gcttcacatttcgctgngtgcattagacaaacttaagctgacgaagaact 685
::: |||||
99 phegluleuthrglnluyasalserysmetcluhphelusalalegluse 115
::: |||||
666 atggctgtactctcgaaagaaacgaaacagaaattttaaactctcggaag 735
::: |||||
115 rvalleuclnhrhlleasprohstrhnglyglulysvalcysleuser 132
::: |||||
736 actctt...actgacacaaacgt...gggaaagactcagcttgact 779
::: |||||
132 ytracyalualasprmetaspraggluleuproalaleumetglualser 148
::: |||||
780 cttaactgacgaatattgacggaatagatcaacttcttgagggtttcc 829
::: |||||
149 lysalavalaleugluasrvalilpherethlsglnasrgluseraant 165
::: |||||
830 aagctgtgcctaaataatgctcattttctgctatcagaagacttttaattg 879
::: |||||
165 proleuclnaspserthrleuylslyspheaspralierhes 182
::: |||||
880 gcccttaagtgaagaaagcgttggaagcaaaagtttgatgacagttttt 929
::: |||||
182 eratstrargtythrlysalaleugluvalilleylslysluethllys 198
::: |||||
930 cagcaacagatnactttaaaccttgaaamaccttgcgacgtcgtcag 979
::: |||||
199 asrglmetglnluilleystrthpherargleuylsleugluasaleu 215
::: |||||
980 acacaaagctcagaaagcttaaaagatatataatggacacttaaatcttcaa 1029
::: |||||
215 nrthralyluasrglnalahllyslleuargluasnlleaglinsp 232
::: |||||
1030 ggaatataagaaaaagcgttgagatctcgatcacagttaacagtaag 1079
::: |||||
232 lnglulyseraspalseryssetrglmetglnleuylsleuglulys 248
::: |||||
1080 aagccacagttacactttcnaagaaatgtccaaatctatgaaacaa 1129
::: |||||
249 ilcysgltthglarggluleuclnmetglnuthrseryleaspgl 265
::: |||||
1130 cttgactccatttaaaahicgtctaaagaaatgtgacataatctctctaa 1179
::: |||||
265 uleuargargleuglnglylnleasprlleyalathrgluargser 282
::: |||||
1180 aatataagaaacttgacaaatgaaatgaaaccccttgagatgccgaagaagac 1229
::: |||||
282 hrleuclnuthrglnlinsgiculylsleualalaleusergluclasn 298
::: |||||
1230 aataggacaaacatnactgacactggaaagaaatggaaaagctttt 1279
::: |||||
299 gluasprhraspgluugluuclutluprlnhrthrsphlegluar 315
::: |||||
1280 caaggaactgagacaaactaaatgacttatatcacaatcaccagacac 1329
::: |||||
315 gllealaleuclnuthrlysllestrlyseuvalargasprmetaspa 332
::: |||||
1330 actgaagcgaagaaagaaatgttgtaactgtctatctatcactgga 1379
::: |||||

332 spglnlsleryserservleuyserservsglnasrservleuthr 348
1380 AACTAATTAATAATCTAGGCTTCAATCGAATAATCTGAAACTGCT 1429
349 Htsglulleglylvleuglnlagnlualasplahstleuthrmely 365
1430 GTTGAACAGGCGTCTACAGCTGCAGACAGATGCCATCAAGAACTAT 1479
365 shstsglulgaasperasplileysasnilecysThrlyshsasnleug 382
1480 CCGAGCTAGAGATTCTAATCAATCACTCTTTGGCAACACAGCTAAGTTGG 1529
382 lYpProvalProclunhsProphethrnsasbvalAlamecasleuthr 398
1530 ATGCGTTTGAGGCTGGACATTCAGTCAAGTGAAGACAGATTAAAAATTTTCAC 1579
399 AsnArglllelyAlaArg.....LeuserSerle 408
1580 AAACCTGTAGAGACAGACAAAGACGGAAGAAACAACTGCCAACCAACT 1629
408 ugluasnaPleuleasPlyleYsSerasnGlnuhsplnleuhsPV 425
1630 GATGATGACTTTCGAGAAAAGACACTCTGAACAAAACAGATAGAT. 1678
425 alleuthrplyshstYrleuyslleasnAlaArgTYserGlnuvalasp 441
1679GAGTAAGAGATMAAATAAGCTGACTGGCA 1708
442 GlylInleGlnserlyslleGlnsermetserGylleLeuArgArgAr 458
1709 AGAATTAATTGAGTTAAATCAATAATCCTAAGT.....AA 1743
458 glysAspLysglulysGluArgasplAlaAlaGlnleuSerlySP 475
1744 GAAGCAGATGAGCTGAAAAAGTGAAGATGAATTAACAGCAGTTGGAG 1793
475 heaslnleuseratgylleasplGluArgGlnuArgHsImecGlnleGlnval 491
1794 GATCTTCACAGAGATTCTTGAACAGCTGCAGACAGAGCTC...ATTAAGCT 1840
492 GluArgLyshThrleuAlaleuGlnylGluArgasPYrAspSerlleIleSe 508
1841 GAACGCTGAG.....TTAAGCAAGCTGCAGAAAACAGCAATGTGA 1881
508 rglnlyAsrGthrglnuvalTYrSerleuglnuhs..... 520
1882 AACCTTAATAATGAGAGTAAATAAGTCTCCAAATGAATAAGACAGACTTGA 1931
521llelysvAlleuleuArgGlnulysAspllelleasnArgasn 534
1932 ACAGAGACCGCTGCTAAACTGTGACAGAGATGAGACAGATTAAACCATAT 1981
535 AlasplGlnuArgVallyserleugly...leuYslyAspAlaLeuGlnSe 550
1982 ACACACACACAGTACCCAAATGAGATGCTGCACCAAGACAAAAGCTGACAA 2031
550 rserlyAspLyseleuasnGlnulevalasnlnuhslyAspLyseIlel 567
2032 AGAT...GAACAATACAGAAAAATAAATCTAGGCACAGTGTGATGAATTA 2078
567 yslYsvAlleuArglyArgasnProPhelGlnulysAspMetlyslu 583
2079 CCTCACAGTTGG...GGATATTTTCCCMACAAAAACAGCTTGACAC... 2122
584 lleasnGlnAlaPheThrProvalAspLysgluTYasnGlnleuArgSe 600
2123TGG.....CTACATAG 2133
600 rlysserGlnGlnAlaGlnGlnleuYsPheThrGlnserlyValT 617
2134 TAAATCAAAAGAAATTAATACAG..... 2155
617 hrAspAlaArgGlnGlnleuThrlysvleuArgArgAspMet..... 630

```

2156 ..... ACCAGGACAGACTTGCCTTAAATTCACAGAGACTTACATCATCT 2200
631 AspAlaLysArgArgPheLeuAspSerLysLeu...GlnSerIleLeuG1 646
::: :::::::::::::::::::: ::::::::::::::::::::
2201 GAGCAGCAATTAATAATCATATAATAATGAACTAAAAAGAAAGAGAGACA 2250
646 nIleSerAlaAsnValAspMetPheProLysValLeu.....GlnAspA 661
::: :::::::::::::::::::: ::::::::::::::::::::
2251 GTTGTCCAGTTCAGAAAGACAGACAGCTGTTCATGTTGTGTGTACCCGAGATT 2300
661 IametaLysArgAspGluGlnLysArgLeuGlnAsnPhaLeuAsnGly 677
::: :::::::::::::::::::: ::::::::::::::::::::
2301 TTGAAATGATTTAGACAGCTTAAAGAGAAATGAAATTAATCAATCAAAA 2350
678 MetArgGluMetLeuAlaPro.....PheG1 686
::: :::::::::::::::::::: ::::::::::::::::::::
2351 CAGCAGAGCAGCTGCTGCTGAGCCACAGACAGTTTACTCCAGTTCATTCAC 2400
686 uHsLeuAlaArgLysAsnHis...ValCysProCysGluArgAlaP 702
::: :::::::::::::::::::: ::::::::::::::::::::
2401 TCAGCTAACAGAGAAACCACTCATGTGCCCCGTTTGTGAGAGAGTTT 2450
702 he...ThrProAspGluGlnAspGluPheValLysGlnArgMetGln 717
::: :::::::::::::::::::: ::::::::::::::::::::
2451 TTCAGACAGAGGCTGAGTTACAGAAAGATCATCAGTGAATTCAGTCTAAA 2500
718 AsnSerSerThrAlaGluArgSerLysAlaLeuAlaMetGluSerSerAs 734
::: :::::::::::::::::::: ::::::::::::::::::::
2501 CTGCGACATTCCTCCACAGATTAACATCAAGTCAACAGATTCAGAGCTAAAA 2550
734 nAlaGluAlaLeuPheGlnGlnLeuAspLysLeuArgThrIleTyAspA 751
::: :::::::::::::::::::: ::::::::::::::::::::
2551 AAAGCAAAAGCGCGCTGATGAATGCTGTGCGACTGTGCCCCAGAGCGAAA 2600
751 IatYValLysLeuValGluGluThrIlePro...LeuAlaGluLys... 765
::: :::::::::::::::::::: ::::::::::::::::::::
2601 GCATTAATGATTGAAAGAGAGAAATACCGAATTAAGAAACCAACCTG 2650
766 ...AspLeuAsnGlnHisLysLeuAlaAspGluSerGlnLysAlaGlnAlaP 781
::: :::::::::::::::::::: ::::::::::::::::::::
2651 CAGATTCGCAATAGACACATACAGCCCTAAAGAACAGACATAGAAACAA 2700
781 eAspAspLeuLeuGly..... 786
::: :::::::::::::::::::: ::::::::::::::::::::
2701 AGAAACACTCTGGGTACATAATAGCTGAGAGAAAGTGGCCAAAGTAT 2750
787 .....ValLeuAlaHisValGlnMetAspArgAla 797
::: :::::::::::::::::::: ::::::::::::::::::::
2751 GCCTGACAGATGTTACAAATTATGAGAGGTTCCAGATGGAATTAAGAT 2800
798 ValGlu.....AlaLeuLeuGlnProThrAspTh 807
::: :::::::::::::::::::: ::::::::::::::::::::
2801 GTTGAAGAAAATTCACACACAGACAGCTAACTCAAGAAATACAC... 2848
807 rIleAspArgHisValHisGlnIleGlnGlnLeuValLysGluValGlnA 824
::: :::::::::::::::::::: ::::::::::::::::::::
2849 TTAGATCGAAGCTGCCAACAGTCAACACAG.....GAGAAACAAAG 2888
824 sPLeuGlnLysArgAlaLeuAspSerSerGlyArgGlyValLysSerLeuGln 840
::: :::::::::::::::::::: ::::::::::::::::::::
2889 AGAAACAGACAAAGTTAGACACA.....GTTTCTAGTAAGATT 2926
841 GlnIleGlnLeuGlnLysArgPheLeuGlnArgThrArgAspThrLeuI 857
::: :::::::::::::::::::: ::::::::::::::::::::
2927 GAATTGAATCGTAAGCTT..... 2944
857 eValGluValAspAspLeuArgAspGlnHisArgMetLeuAsnGluAsp 874
::: :::::::::::::::::::: ::::::::::::::::::::
2945 .....ATACAGACACAGAGAAACAGATTCACACATCTAA 2978
874 eSerSerAlaGlnValArgThrHisAsnAlaArgGluGluLysValLys 890
::: :::::::::::::::::::: ::::::::::::::::::::
2979 AAAGTACACA.....AATGACCTAAATTCGAGAACTTCAG 3016
891 AlaSerSerIleLeuGlnArgPheGlnLysSerGluGlnGluValLe 907
::: :::::::::::::::::::: ::::::::::::::::::::
3017 AATATCCATTAATTTGCAACGTGCTCAGCAACTGAGGAGACACAGCTGGA 3066
907 uLeuAlaGlnGluLysGlnGlnLeuLeuValGlnLysLysLeuLeuGlnG 924
::: :::::::::::::::::::: ::::::::::::::::::::
3067 ATTATCCACTGAAGTTCAGTCTTTGTACAGAGAGATTAAGATGCTTAAG 3116
924 IuSerLeuAspProLeu.....SerLysGluLys 933
::: :::::::::::::::::::: ::::::::::::::::::::
3117 ACAGGTAAAGCCCTTTGAAACACACATGGAAGATTCACAGACAGAAAA 3166
934 GluSerLeuLeuGlnGlnLysArgAlaLeuLysGlnLysLeuAspGluG 950
::: :::::::::::::::::::: ::::::::::::::::::::
3167 GAAGATTAATCAACAAAAAATTAACACCAACAAATAGCAGCAAGTAA 3216
950 uTyHisGlnLeuAlaGlnArgLysArgGluPheGlnGlnLysLeuAspA 967
::: :::::::::::::::::::: ::::::::::::::::::::
3217 ACTGAATGATTAATAAGAGAGGTTAAATATCTATGCTATATG... 3262
967 IalLeuGlnArgLeuAsnMetLysIleLysGlyTyLeuAspSerLysLys 983
::: :::::::::::::::::::: ::::::::::::::::::::
3263 .....AAAGACATTCAGAAATTAATTCAGATGGGAAA 3295
984 AsnGlu.....LysLeuLysGluLeuGlnGlnLysValLeuCysH1 998
::: :::::::::::::::::::: ::::::::::::::::::::
3296 GACAGCTAATAGAACCAAAAGAACTGAATTAATTAATTA... 3340
998 sSerGlnLeuGlnSerCysMetAlaLysGlnGlnArgIleSerAlaGlnL 1015
::: :::::::::::::::::::: ::::::::::::::::::::
3341 .GCTCACTAAGTAATGCGAGAAACACAAAGAAAGATTAATTAAGATA 3389
1015 euAsnLysSerLysGlnLeuLeuGlnGlnGlnLysLysArgAsn 1031
::: :::::::::::::::::::: ::::::::::::::::::::
3390 TCGAGACTCATGACACAAAGATTTGATACACAGAAAGATACAAAGAGCTG 3439
1032 rIleAspAspAsnLeuLysTyArgLysThrLysAlaAspValGluGlnLe 1048
::: :::::::::::::::::::: ::::::::::::::::::::
3440 CTACAGATTAACCTTACTTAAGAAAGAAATGAGAACTTAAGAAAGAACT 3489
1048 uThrArgAspIleGluSerLeuGlnGlnArgLeuLeuSerIleGlySerL 1065
::: :::::::::::::::::::: ::::::::::::::::::::
3490 TGA...GAGAAAGAAACACACATTTGAAGAAATGGCGAAA 3530
1065 euSerAlaIleGlnAlaAspLeuLysArgHisSerGlnLys... 1079
::: :::::::::::::::::::: ::::::::::::::::::::
3531 TCGAGCTTTGCCAAATGAAGAAAGTGAACATCAGAACTGGAGAGACATA 3580
1080 GluArgLeuAsnSerGluPheAsnArgTyArgGlnIleThrLeuSerValTy 1096
::: :::::::::::::::::::: ::::::::::::::::::::
3581 GACAAATATAAAGAAATCATATAATTGGCATTTGGCGCAGACAGAAAGTTA 3630
1096 rGlnSerAsnIleSerLysHisLysGlnGlnLeuLysLeuSerGlnTyL 1113
::: :::::::::::::::::::: ::::::::::::::::::::
3631 TGAAGAGAAATTAATTCATTTAAGAAAGAACTCGAAGAACCAATTTTC 3680
1113 yAspIleGlnLysArgTyThrAsnGlnPheLeuGlnLeuLysThrTh 1129
::: :::::::::::::::::::: ::::::::::::::::::::
3681 GCGATCTCGAGAAAGATAGAGAAATGATGATTTTATGAGAGACACAA 3730
1130 GluMetAlaAsnLysAspLeuAspArgTyTyThrAlaLeuAspLysAl 1146
::: :::::::::::::::::::: ::::::::::::::::::::
3731 GAACCTGTGAACAAGATCGATATTTAATAAGACCTTGACCAAC 3780
1146 AlaMetArgPheHisSerMetLysMetGlnGlnLysLeuLysIleLeu 1163
::: :::::::::::::::::::: ::::::::::::::::::::
3781 AATTAATGAATTTCAACAGATGAAGAAATGGAAGAAATCAATTAATATAC 3830
1163 yGlnLeuLeuTyArgGlnThrTyArgGlyGlnAspIleAspTyIleSer 1179
::: :::::::::::::::::::: ::::::::::::::::::::
3831 GTGACCTGTGCGAAGTACCTATCTGAGACAAATATGATATCAATAGAA 3880

```

```

1180 IleasnSerasp.....SerGluGlyAlaGlyThrArgys 1191
|||||
3681 ATACGGTCTATGCCGATGAATAATGATACGCTTGATTAAGGCGGAA 3930
|||||
1191 RTYSerThrArgValAlaMetGlnThrGlyAspAlaGluLeuGluMetA 1208
|||||
3931 TTATACCTACCGAGTGTGATGCTGGAAGGAGACACACCTTGATATGC 3980
|||||
1208 TGGTATGTCYSerAlaGlyGlnLysValLeuAlaSerLeuIleIleArg 1224
|||||
3981 GAGGACGATGCAAGTGTGACAAAGGATTAAGCTCCTCATCATTCGCC 4030
|||||
1225 LeuAlaLeuAlaGluThrPheCysLeuAsnCysGlyIleLeuAlaLeuAs 1241
|||||
4031 CTGCGCCCTGGCTGAAGCTTCTGCTCACTGAGCATTCATGCTTGGA 4080
|||||
1241 PGluProThrThrAsnLeuAspGlyProAsnAlaGluSerLeuAlaIle 1258
|||||
4081 TGACCCACACAAATCTTGACGAGAAACATGTAATCTCTCCACATG 4130
|||||
1258 IaleuLeuArgIleMetGluAlaAlaArgLysGlyGlnGluAsnPheGlnLeu 1274
|||||
4131 CTCGTGTGAGATATAATAAGTCGTACAGCAGCGTAATCTCCACCTT 4180
|||||
1275 ILevalIleThrHisAspGluArgPheAlaHisLeuIleGlyGlnArgL 1291
|||||
4181 CTGCTAATCACTCATGATGAGATTTGTGAGCTTTAGAGCTTCTGA 4230
|||||
1291 nLeuAlaGluLysTyrtYrArgValSerLysAspGluAsnGlnHisSerI 1308
|||||
4231 ATATGTGAGAAATCTACAGGATTAATAAGACATGATCATGCTCAG 4280
|||||
1308 IleIle 1309
|||||
4281 AGATT 4285

```

seq_name: /cgn2.6/prodata/2/lna/5A.COMB.seq:US-08-592-126-54

seq_documentation_block:

Sequence 54, Application US/08592126

Patent No. 5821091

GENERAL INFORMATION:

APPLICANT: Gregory Dolganov

TITLE OF INVENTION: Transcripts Encoding Immunomodulatory

TITLE OF INVENTION: Polypeptides

NUMBER OF SEQUENCES: 151

CORRESPONDENCE ADDRESS:

ADDRESSEE: Dehlinger & Associates

STREET: 350 Cambridge Avenue, Suite 250

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94306

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/592,126

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Shultz, Charles K.

REGISTRATION NUMBER: 38,615

REFERENCE/DOCKET NUMBER: 4600-0111

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 324-0880

TELEFAX: (415) 324-0960

INFORMATION FOR SEQ ID NO: 54:

SEQUENCE CHARACTERISTICS:

LENGTH: 5893 base pairs

```

? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear
? MOLECULE TYPE: cDNA to mRNA
? HYPOTHETICAL: NO
? ANTI-SENSE: NO
? ORIGINAL SOURCE:
? INDIVIDUAL ISOLATE: Rad50.seq
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 389..4324
? US-08-592-126-54

```

```

alignment_scores:
  Quality: 1729.00      Length: 1385
  Ratio: 1.919          Gaps: 38
  Percent Similarity: 65.054  Percent Identity: 32.419

```

alignment_block:

US-09-538-396-2 x US-08-592-126-54

Align seq 1/1 to: US-08-592-126-54 from: 1 to: 5893

```

1 MetSerThrValAspLysMetLeuIleLysGlyIleArgSerPhe..... 15
|||||
389 ATGCCCGGATCCGAAAGATGAGCATTCGTGGCGGTGGGAGTTTGAT 438
|||||
16 AspProAspAsnLysAsnValIleThrPhePheLysProLeuThrLeuI 32
|||||
439 AGAGACAAAGATATGCAATTAATCACTTCTTCAGCCCTTACAAATT 488
|||||
32 LevalGlyProAsnGlnAlaGlyLysThrThrIleIleGluCysLeuLys 48
|||||
489 TGGTGGACCCAAATGGGGGAGGAGAGACCATCATTCATGATGCTAATA 538
|||||
49 LeuSerCysThrGlyGluLeuProProAsnSerArgSerGlyHisThrPh 65
|||||
539 TATATTGTACTGAGATTTCCCTCGGAACCAA...GGAATACATT 585
|||||
65 eValHisAspProLysValAlaGlyGluThrGluThrLysGlyGlnIleL 82
|||||
586 TGTACAGATCCCAAGGTGCTCAAGAAACAGATGAGACCCAGATTC 635
|||||
82 ySLeuAlaGpHeLysThrAlaAlaGlyLysAspValValCysIleArgSer 98
|||||
636 GTCGCAATTTCGTATGTCATGAGAGAACTTATAGCTGTCAGAAATCT 685
|||||
99 PheGlnLeuThrGlnLysAlaSerLysMetGluPheLysAlaIleGluSe 115
|||||
686 ATGTGTGTACTCGAAGAAAGCAAAAGACAGAAATTAATAACTCGGAAG 735
|||||
736 AGTCATT...ACTAGAACAAAGCAT...GGTGAAGAGTCAGCTGAGCT 779
|||||
115 rValLeuGlnThrIleAsnProHisThrGlyLysValCysLeuSerT 132
|||||
132 YrArgCysAlaAspMetAspArgGluIleProAlaLeuMetGlyValSer 148
|||||
780 CTAGTGTGCGAAGAAATTCACCGAAGAAATGATCAGTCTCTGGGGTTCC 829
|||||
149 LysAlaValLeuGluAsnValIlePheValHisGlnAspGluSerAsnTr 165
|||||
830 AAGGTGTGCTAATATATGTCATTTTCTGTCATCAAGAAATTCATATG 879
|||||
165 pProLeuGlnAspProSerThrLeuLysLysLysPheAspIlePheS 182
|||||
880 GCCTTAAGTGAAGAAAGCTTGAAGCAAAAGTTGATGATATTTT 929
|||||
182 eValIleThrArgTyThrLysAlaLeuGluValIleLysLysLeuHisLys 198
|||||
930 CAGCAACAAAGATACATTAAAGCCCTTAAGCACTTGGCAGGTAGTGG 979
|||||
199 AspGlnMetGlnGluIleLysThrPheArgLeuLysLeuGluAsnLeuG 215

```

111
980 ACACAAAGCAGAACTAAAGAAATATCAATGGAACCTAAATATCTGAA 1029
215 nThValLysArgGlnAlaHisLysLeuArgGlnAsnIleAlaGlnAsp 232
1030 GCATATGAAAGAAAGCTTGTGAGATCTGATCAGATTACAACTAAAG 1079
232 InGluLysSerAspAlaSerLysSerGlnMetGlnInLeuLysGlu 248
1080 AAGCCCAATTAACATCTTCAAGGAATTCATAATCTATGAGAAATGAA 1129
249 IleCysGluThrGluArgGluIleLeuGlnMetGluThrSerLeuAsp 265
1130 CTGTGATCCATTGACAAATCTCTAAAGAAATGAAATATATCTCTTA 1179
265 uLeuArgArgLeuGlnGlnIleAspIleLysAlaThrGluArgSer 282
1180 AATTAATGAACCTTGACAAATGAATTAAGCTTGATAGCCGAAGAGC 1229
282 hLeuLeuThrGlnGlnHisGluLysLeuAlaLeuSerGluGlnAsn 298
1230 AATGAGAGAAATATATAGTGAACCTGGAAGAGAAATGGAAGCTTTT 1279
299 GluAspThrAspGluGluLeuMetGluTrpGlnThrLysPheGluArg 315
1280 CAAGGAGCTGATGACAACTAAATGACTTATATCACAATCACCAGAAC 1329
315 gIleAlaLeuLeuGluThrLysIleSerLysLeuValArgAspMetAsp 332
1330 AGTAAGGAGAAAGAAAGAAATGTGTAGACTCTCATCTGAACTGGAA 1379
332 spGluLysSerLysSerValLeuSerLysGlnAsnSerGluLeuThr 348
1380 AACTAATATGAATCTAGGCTTCTCAATCAGGAAATTCAGAACTGCTT 1429
349 HisGluIleGluLysLeuGlnAlaGluAlaAspAlaHisLeuThrMet 365
1430 GTTGAACAGGCTGCTCTACAGCTGCAAGCAGATGCCATCAGAACAT 1479
365 sHisGluArgAspSerAspIleLysAsnIleCysThrLysHisAsnLeu 382
1480 CCGAGCTAGAGATTCATTAATCAGCTTGGCACAACAGCTGAATGG 1529
382 IyrProValProGluHisProPheThrAsnAspValAlaMetAsnLeuThr 398
1530 ATGGCTTGAGCGTGGCCATTCAGTGAAGACAGATTAATAATTTCAC 1579
399 AsnArgIleLysAlaArg.....LeuSerSerIle 408
1580 AAACCTGTAGAGAGAGACAGAAAGGGAACAAAACTGCCAACCACT 1629
408 uGluAsnAspLeuLeuAspLysLysLysSerAsnGluAspGlnLeuAsp 425
1630 GATGAATGACTTTCAGAAAAAGAGACTGGAACAAACAAACACATGAT 1678
425 alLeuThrLysHisLysLysIleAsnAlaArgTyrSerGluValAsp 441
1679GAGATAAGAGATGAAGAAACTGAGCTGGA 1708
442 GlyGlnIleGlnSerLysIleGluSerMetSerGlyIleLeuArgArg 458
1709 AGAATATGACTTAAATCAGAAATCCTAAGT.....AA 1743
458 gLysAspLysGluLysGluArgAspAlaAlaGluValGluLeuSerLys 475
1744 GAGCGCAAAAGAGCTGAAAAATGTGAAGTATGAAATTCACAGCTGGAG 1793
475 hAsnLeuSerArgIleAspGluArgGluArgHisMetGlnIleGluVal 491
1794 GATCTTCAGACAGATCTTGAACCTGACAGAGAGCTC...ATTAAGCT 1840
492 GluArgLysThrLeuAlaLeuGluGluArgAspTyrAspSerIleIle 508

1841 GAACGTGAG.....TTAAGCAAGCTGAGAAAAACGCAATGTAGA 1881
508 rGlnLysArgThrGluValLysSerLeuGluGlnLys..... 520
1882 AACCTTAATAATGAAATGATTAATGCTCCAAATGAAAGAGAGATTAG 1931
521IleLysValLeuLeuArgGluLysAspIleIleAsnArgAsn 534
1932 ACAGAGACCTGCTTAACCTGACAGAGAGATGAGAGAGTTAAACATCAT 1981
535 AlaAspGluArgValLysLeuGly...LeuLysLysAspAlaLeuGlu 550
1982 ACAACAACAGCTACCAAAATGAGATGCTGACCAAGACAAACCTACAA 2031
550 rSerLysAspLysLeuAsnGluIleValAsnGluHisLysAspLysIle 567
2032 AGAT...GAACAATCAAGAAAAATTAATTTAGCCACAGTGAATTA 2078
567 yLysValLeuArgGluArgAsnProPheGluLysAspMetLysGlu 583
2079 CCTCACTGTG...GGATATTTCCACAAACAAAGAGCTGAAGAC... 2122
584 IleAsnGlnAlaPheTrpProValAspLysGluTyrAsnGluLeuArg 600
2123TGG.....CTACATAG 2133
600 rLysSerGlnGluAlaGluGlnLeuLysPheThrGlnSerLysVal 617
2134 TAAATCAAAAGAAATTAATCAG..... 2155
617 hAspAlaArgGluGlnLeuThrLysLeuArgArgAspMet..... 630
2156ACCAAGGAGAGCTTGCCAAATTTGACAAAGAGAGCTTATCT 2200
631 AspAlaLysArgArgPheLeuAspSerLysLeu...GlnSerIleLeuG 646
2201 GAGCAGAAATTAATAATCATTAATGAATCAATAAGAGGAGAGCA 2250
646 nIleSerAlaAsnValAspMetPheProLysValLeu.....GlnAsp 661
2251 GTTGTCCAGTTAGCAAGACAGCTGTTGATGTTGTGTGAGCCAGGAT 2300
661 lMetAsnLysArgAspGluGlnLysArgLeuGlnAsnPheAlaAsnGly 677
2301 TTGAAGTGAATTAAGACAGCTTAAGAGAAATGGAATAATCATCAAA 2350
678 MetArgLysLeuAlaPro.....PheG 686
2351 CAGGAGACCATGCTGCTGAGACGACAGCAAGTTTACTCCAGTTCAATTAC 2400
686 uHisLeuAlaArgLysAsnHis...ValCysProCysGlyGluArgAla 702
2401 TCAGCTAACAGACAAACAGATCATGTTGCTGCTGTCAGAGAGATTT 2450
702 he...ThrProAspGluGluAspGluPheValLysLysGlnArgMetGln 717
2451 TTCAGACAGAGGCTGAGTTACAAAGAGCTCATGATTTGCACTGCTAAA 2500
718 AsnSerSerThrAlaArgSerLysAlaLeuAlaMetGluSerSerAs 734
2501 CTGCGACTTGTCTCAGATTAACCTCAACTCAACAGAAATCAGAGCTAAAA 2550
734 nAlaGluAlaLeuPheGlnGlnLeuAspLysLysLeuArgThrIleTyrAsp 751
2551 AAAGAAAGAGCGGCTGATGAAATGCTGGAGCTGTGCCATGAGCGCA 2600
751 lArgValLysLeuValGluGluThrIlePro...LeuAlaGluLys... 765
2601 GCATTAATGATTTGAAGAGAGAAATACCAATTAAGAAACAAACCTG 2650
766 ...AsnLeuAsnGlnHisLeuAlaAspGluSerGlnLysAlaGlnAlaPhe 781
2651 CAGAAATGTCATTAAGACATACAGCGCTAAAGAAAGACATAGAAGACA 2700

```

781 eaaspleuengly..... 786
782 .....
783 .....
784 .....
785 .....
786 .....
787 .....
788 .....
789 .....
790 .....
791 .....
792 .....
793 .....
794 .....
795 .....
796 .....
797 .....
798 .....
799 .....
800 .....
801 .....
802 .....
803 .....
804 .....
805 .....
806 .....
807 .....
808 .....
809 .....
810 .....
811 .....
812 .....
813 .....
814 .....
815 .....
816 .....
817 .....
818 .....
819 .....
820 .....
821 .....
822 .....
823 .....
824 .....
825 .....
826 .....
827 .....
828 .....
829 .....
830 .....
831 .....
832 .....
833 .....
834 .....
835 .....
836 .....
837 .....
838 .....
839 .....
840 .....
841 .....
842 .....
843 .....
844 .....
845 .....
846 .....
847 .....
848 .....
849 .....
850 .....
851 .....
852 .....
853 .....
854 .....
855 .....
856 .....
857 .....
858 .....
859 .....
860 .....
861 .....
862 .....
863 .....
864 .....
865 .....
866 .....
867 .....
868 .....
869 .....
870 .....
871 .....
872 .....
873 .....
874 .....
875 .....
876 .....
877 .....
878 .....
879 .....
880 .....
881 .....
882 .....
883 .....
884 .....
885 .....
886 .....
887 .....
888 .....
889 .....
890 .....
891 .....
892 .....
893 .....
894 .....
895 .....
896 .....
897 .....
898 .....
899 .....
900 .....
901 .....
902 .....
903 .....
904 .....
905 .....
906 .....
907 .....
908 .....
909 .....
910 .....
911 .....
912 .....
913 .....
914 .....
915 .....
916 .....
917 .....
918 .....
919 .....
920 .....
921 .....
922 .....
923 .....
924 .....
925 .....
926 .....
927 .....
928 .....
929 .....
930 .....
931 .....
932 .....
933 .....
934 .....
935 .....
936 .....
937 .....
938 .....
939 .....
940 .....
941 .....
942 .....
943 .....
944 .....
945 .....
946 .....
947 .....
948 .....
949 .....
950 .....
951 .....
952 .....
953 .....
954 .....
955 .....
956 .....
957 .....
958 .....
959 .....
960 .....
961 .....
962 .....
963 .....
964 .....
965 .....
966 .....
967 .....
968 .....
969 .....
970 .....
971 .....
972 .....
973 .....
974 .....
975 .....
976 .....
977 .....
978 .....
979 .....
980 .....
981 .....
982 .....
983 .....
984 .....
985 .....
986 .....
987 .....
988 .....
989 .....
990 .....
991 .....
992 .....
993 .....
994 .....
995 .....
996 .....
997 .....
998 .....
999 .....

```

```

1048 utHArgAspIleGluSerLeuGluArgLeuSerIleGlySerL 1065
1049 .....
1050 .....
1051 .....
1052 .....
1053 .....
1054 .....
1055 .....
1056 .....
1057 .....
1058 .....
1059 .....
1060 .....
1061 .....
1062 .....
1063 .....
1064 .....
1065 .....
1066 .....
1067 .....
1068 .....
1069 .....
1070 .....
1071 .....
1072 .....
1073 .....
1074 .....
1075 .....
1076 .....
1077 .....
1078 .....
1079 .....
1080 .....
1081 .....
1082 .....
1083 .....
1084 .....
1085 .....
1086 .....
1087 .....
1088 .....
1089 .....
1090 .....
1091 .....
1092 .....
1093 .....
1094 .....
1095 .....
1096 .....
1097 .....
1098 .....
1099 .....
1100 .....
1101 .....
1102 .....
1103 .....
1104 .....
1105 .....
1106 .....
1107 .....
1108 .....
1109 .....
1110 .....
1111 .....
1112 .....
1113 .....
1114 .....
1115 .....
1116 .....
1117 .....
1118 .....
1119 .....
1120 .....
1121 .....
1122 .....
1123 .....
1124 .....
1125 .....
1126 .....
1127 .....
1128 .....
1129 .....
1130 .....
1131 .....
1132 .....
1133 .....
1134 .....
1135 .....
1136 .....
1137 .....
1138 .....
1139 .....
1140 .....
1141 .....
1142 .....
1143 .....
1144 .....
1145 .....
1146 .....
1147 .....
1148 .....
1149 .....
1150 .....
1151 .....
1152 .....
1153 .....
1154 .....
1155 .....
1156 .....
1157 .....
1158 .....
1159 .....
1160 .....
1161 .....
1162 .....
1163 .....
1164 .....
1165 .....
1166 .....
1167 .....
1168 .....
1169 .....
1170 .....
1171 .....
1172 .....
1173 .....
1174 .....
1175 .....
1176 .....
1177 .....
1178 .....
1179 .....
1180 .....
1181 .....
1182 .....
1183 .....
1184 .....
1185 .....
1186 .....
1187 .....
1188 .....
1189 .....
1190 .....
1191 .....
1192 .....
1193 .....
1194 .....
1195 .....
1196 .....
1197 .....
1198 .....
1199 .....
1200 .....
1201 .....
1202 .....
1203 .....
1204 .....
1205 .....
1206 .....
1207 .....
1208 .....
1209 .....
1210 .....
1211 .....
1212 .....
1213 .....
1214 .....
1215 .....
1216 .....
1217 .....
1218 .....
1219 .....
1220 .....
1221 .....
1222 .....
1223 .....
1224 .....
1225 .....
1226 .....
1227 .....
1228 .....
1229 .....
1230 .....
1231 .....
1232 .....
1233 .....
1234 .....
1235 .....
1236 .....
1237 .....
1238 .....
1239 .....
1240 .....
1241 .....
1242 .....
1243 .....
1244 .....
1245 .....
1246 .....
1247 .....
1248 .....
1249 .....
1250 .....
1251 .....
1252 .....
1253 .....
1254 .....
1255 .....
1256 .....
1257 .....
1258 .....
1259 .....
1260 .....
1261 .....
1262 .....
1263 .....
1264 .....
1265 .....
1266 .....
1267 .....
1268 .....
1269 .....
1270 .....
1271 .....
1272 .....
1273 .....
1274 .....
1275 .....
1276 .....
1277 .....
1278 .....
1279 .....
1280 .....
1281 .....
1282 .....
1283 .....
1284 .....
1285 .....
1286 .....
1287 .....
1288 .....
1289 .....
1290 .....
1291 .....
1292 .....
1293 .....
1294 .....
1295 .....
1296 .....
1297 .....
1298 .....
1299 .....
1300 .....
1301 .....
1302 .....
1303 .....
1304 .....
1305 .....
1306 .....
1307 .....
1308 .....
1309 .....
1310 .....
1311 .....
1312 .....
1313 .....
1314 .....
1315 .....
1316 .....
1317 .....
1318 .....
1319 .....
1320 .....
1321 .....
1322 .....
1323 .....
1324 .....
1325 .....
1326 .....
1327 .....
1328 .....
1329 .....
1330 .....
1331 .....
1332 .....
1333 .....
1334 .....
1335 .....
1336 .....
1337 .....
1338 .....
1339 .....
1340 .....
1341 .....
1342 .....
1343 .....
1344 .....
1345 .....
1346 .....
1347 .....
1348 .....
1349 .....
1350 .....
1351 .....
1352 .....
1353 .....
1354 .....
1355 .....
1356 .....
1357 .....
1358 .....
1359 .....
1360 .....
1361 .....
1362 .....
1363 .....
1364 .....
1365 .....
1366 .....
1367 .....
1368 .....
1369 .....
1370 .....
1371 .....
1372 .....
1373 .....
1374 .....
1375 .....
1376 .....
1377 .....
1378 .....
1379 .....
1380 .....
1381 .....
1382 .....
1383 .....
1384 .....
1385 .....
1386 .....
1387 .....
1388 .....
1389 .....
1390 .....
1391 .....
1392 .....
1393 .....
1394 .....
1395 .....
1396 .....
1397 .....
1398 .....
1399 .....
1400 .....
1401 .....
1402 .....
1403 .....
1404 .....
1405 .....
1406 .....
1407 .....
1408 .....
1409 .....
1410 .....
1411 .....
1412 .....
1413 .....
1414 .....
1415 .....
1416 .....
1417 .....
1418 .....
1419 .....
1420 .....
1421 .....
1422 .....
1423 .....
1424 .....
1425 .....
1426 .....
1427 .....
1428 .....
1429 .....
1430 .....
1431 .....
1432 .....
1433 .....
1434 .....
1435 .....
1436 .....
1437 .....
1438 .....
1439 .....
1440 .....
1441 .....
1442 .....
1443 .....
1444 .....
1445 .....
1446 .....
1447 .....
1448 .....
1449 .....
1450 .....
1451 .....
1452 .....
1453 .....
1454 .....
1455 .....
1456 .....
1457 .....
1458 .....
1459 .....
1460 .....
1461 .....
1462 .....
1463 .....
1464 .....
1465 .....
1466 .....
1467 .....
1468 .....
1469 .....
1470 .....
1471 .....
1472 .....
1473 .....
1474 .....
1475 .....
1476 .....
1477 .....
1478 .....
1479 .....
1480 .....
1481 .....
1482 .....
1483 .....
1484 .....
1485 .....
1486 .....
1487 .....
1488 .....
1489 .....
1490 .....
1491 .....
1492 .....
1493 .....
1494 .....
1495 .....
1496 .....
1497 .....
1498 .....
1499 .....
1500 .....

```

seq_name: /cgn2_6/prodata/2/lna/5a_COMB.seq:US-08-328-254-5

seq_documentation_block:

; Sequence 5, Application US/08328254

; Patent No. 5710022

GENERAL INFORMATION:

APPLICANT: Zhu, Xueliang
APPLICANT: Lee, Wen-Hwa
TITLE OF INVENTION: A No. 5710022e1 Nuclear Mitotic Phosphoprotein
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/328,254
FILING DATE: 24-OCT-1994
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/141,239
FILING DATE: 22-OCT-1993
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-CU 1191
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 8789 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 544..7990
US-08-328-254-5

alignment_scores:

Quality: 439.50 Length: 1256
Ratio: 0.668 Gaps: 49
Percent Similarity: 52.389 Percent Identity: 19.904

alignment_block:

US-09-538-396-2 x US-08-328-254-5 ..

Align seg 1/1 to: US-08-328-254-5 from: 1 to: 8789

51 CysThrGluGluLeuProCAsnSerArgLysIsthrPheValH1 67
|||||
3808 TGCCTGGGAAACAGTCCCGATACCAAT.....TA 3839
67 SASPrOLysValAlaGluGluThrGluThrLysGluGlnIleLysLeuA 84
|||||
3840 TGACCCCT.....CCAGGGAGAAATAAACCAGGGCTTCAGAA.... 3879
84 rgrPheLysThrAlaAlaGluLysAspValValCysIleArg.....Ser 98
|||||
3880TGCAATTTCGAATTGCA 3897
99 PheGlnLeuThrGlnLysAlaSerLysMetLuphe.....Lys 111
|||||
3898 TTTTCTGGTCTTAATGCTTGTGATGATTCCTCGGGAATCAGCA 3947
111 SAAlIleGluSerValLeuGlnThrIleAsnProHisThrGluGluLysV 128
|||||
3948 AGATATCATTAATCTCAACTGCGGGTAAAGAGACATCAATGAATGAAAT 3997

128 aLysLeuSerTyrArgCysAlaAspMetAspArgGluIleProAlaLeu 144
::: |||
3998 TGAGTATCTTCATCTGATAGAGACCGTGCACAGAAAGTTGAAGTTG 4047
145 MetGlyVal.....SerLysAlaValLeuGluAsnValI1 156
|||||
4048 CTAAATGAATGAAGAATTAAGTACCTCAAACTCCATTACAGAGGTCA 4097
156 ePheValHisGlnAspGluSerAsnTrpProLeuGlnAsp...ProSer 172
|||||
4098 ACTATGACCAAAATTTGAAGCATGATGAAATTTGAAAAAATAGTTGGG 4147
172 hrLeuLysLysPheAspAspIlePheSerAlaThrArgTyrThrLys 188
|||||
4148 AACTTAAGAAAGAAACATCAGATTTA.....AGTGA 4179
189 AlaLeuGluValIleLysLysLeuHisLysAspGlnMetGlnIleLys 205
|||||
4180 AAATGGAAATATTTCTTGTGATCACCAGAGTACTCCAGAGATGA 4229
205 sthrPheArgLeuLysLeuGluAsnLeuGlnThrValLysAspGlnAlaH 222
|||||
4220 AACTTGTGAAGCCCTCAATTCGATTGAAATGATCAGATTAATCAT 4279
222 IsLysLeuArgGluAsnIleAlaGlnAspGlnGluLysSerAspAlaSer 238
|||||
4280 CA.....CGTAAGATATGAGATTAATGCGCCAAAGTGAATGACAGC 4323
239LysSerGlnMetGluGlnLeuLysG1 247
4324 TGGAAAGAGATTTCTGATGTGAAATAGCTGCAATGATGATGATGATC 4373
247 uLysIleCysGluThrGluArgGluIleLeuGlnMetGluThrSerLeuA 264
|||||
4374 GGAGAAAGCTAGCATTTGAGCATGAAGCCCTCACTGAGAGCTGACTAG 4423
264 sPGLuLeuArgArg..... 268
4424 AGGTATGTTCAACAGAGAACTATGTTAGAAAAAGACATGAATGAAG 4473
269LeuGlnGluGlnIleAspIleLysAlaThrG1 279
4474 CAGAAAGTTATTTCTGCTGCTGAAAGAGAGCTCTGATGATGATGATG 4523
279 uArgSerThrLeu.....LeuThrGlnGlnHisGlu 290
|||||
4524 GAGAAACCAAGCTCGTGGAGAAATTAAGATATGTCATAAAAAACACAG 4573
290 yLeuAlaAlaLeuSerGluGluAsnGluAspThrAspGluLeuMet 306
|||||
4574 CACTGATCAGCTTGTCTGAAAAAATGAAGAGAAACAAAGAGCTTAG 4623
307 GluTrpGlnThrLysPheGluGluArgIleAlaLeuLeuGluThrLysI1 323
|||||
4624 TCTCATCAAGAGAGTGTCTCATTCATTCAGGTGCGAGAGGCGAGAGT 4673
323 eSerLysLeuValArgAspMetAspAspGluAlaSerTyrSerSerVal1 340
|||||
4674 GAAGGAAAGAGAGAGAGCTTCACAGACTTTGCTCCCTATGTGATGAGC 4723
340 eSerLysGlnAsnSerGluLeuThrHisGluIleGlyLysLeuGlnAla 356
|||||
4724 TGTTAAGACAAACATCATCTCCAGAAAGAGCTCAGAGTTTGAAGAAG 4773
357 GluAlaAspAlaHisLeuThrMetLysHisGluArgAspSerAspIleLys 373
|||||
4774 GACTCAGAGGCGCTGTTTGAACAATAATGAGTGAAGAAACCAATGTC 4823
373 sAsnIleCysThrLysHisAsnLeuGlyProValProGluHisProPheT 390
|||||
4824 ACACAGTAATAAGAG..... 4839

390 hransapvalalalmetasnleuthrasnarlylleysalaargleus 406
4840 ..AAGGATTGCTTGTCAAGGAATCTGAAGCCTGCGAGGCCAGAGTAGT 4887
407 Serleuglnuansapleuasnleuaplyslylseseransngluaspgle 423
4888 GAATCAGATATATGAAGGCTGAATGTCTCCAGGCCCTTGAGGCCGACT 4937
423 uasapvalleutrpshistyrleuyslleasnalaargtysergluv 440
4938 GGTGAGGAAGAGTGCATTCATTCAGGCTGAGCTCCAAACAGAGGAG 4987
440 alsapgllylnleuglnserlylleuglnserlylleuglnarg 456
4988 TGCAT...CAGCTGAGAGAGGCTGAGAACTGAGA.....GTTCGC 5028
457 Argtrgysaplyslysgluargaspalaalaglulvleuse 473
5029 ATTGAGGCCATGAAGAGAGAGGCTGCACATGCGAG... 5067
473 rlyspheasnleuserarglleaspgluararghismecglnleg 490
5068AACTGAAGAGAGCGGAGCGG..... 5088
490 luvalgluarlysthrleualaleugllyluargasptryraspserile 506
5089GAGAAATGAT..... 5097
507 lleserlnysartrgthrgluvaltyrserleuglnleuyslleysva 523
5098TCACTTAAGATTAAGTTGAGAA 5120
523 lleuueargglulysapllelleasnarganalaaspgluargval 540
5121 CCTGAAGAGGAATTCAGATGTCAGAGAAAGAAC...CAGAGCTAAGTGA 5167
540 yslleugllyleuyslyaspalaaleuglnserlyspaslyleuasn 556
5168 TTCTT.....GATGCC...GAGAAATTCAGAAAGCAAGAGTGAAG 5202
557 glullevaaleuglnuhsleuaplyslylleysalleuarglylar 573
5203 ACCTTA.....AAACACAAATAGAAAGAGTGCAGAGAGCCT 5240
573 gasnprophleuglnuhsapmetlystysglulleasnlnalaphetrrp 590
5241 GAAAGTTTGAATTAACCTTGC..... 5265
590 rovalasplysglutyraanglnleuargserlysserlnglualeglu 606
5366ACGTTAAGGCTCTGMAAAAGAAATCTGACA 5295
607 glnglnleuyspethrlnserlyllevalthrspalaargluginle 623
5396 AAACAAATACAGAAACAAAGT.....CAGTT 5324
623 utrllyleuargtrgshapmetaspalaalysargargphelueaspserl 640
5325 GTCAGAACTGACAAAGTACTCTCTCATTTAAAGCTGTAGAAAGAA 5374
640 yslleuglnserlleleuglnlleaserlalsanvalaspmetpheprolys 656
5375 AGGAGCAGAGAGATACAGATC..... 5397
657 valleuglnaspalametasnleuargaspgluuglnlystargleugluas 673
5398AAGAGAAATCTTAACCTGCAGTGCAGAT 5426
673 npheliasnnglymetargglumetleualaprophleuglnhileualla 690
5427 GCTTCAGAAATGAGTAAGAGCTAAATGAGGAGTAGAGGCTGTGTG 5476
690 rgllysaanhilavalcyaprocycygluaralaphetrrproaspglu 706

5477 GTGACCAAGAAATTAAGAGGCCACAGACAGAGTACAGCCACCAATA 5526
707 gluaspgluphevallystysgluarargmetglnasnerthralagl 723
5527 GAGCAAGAG.....CATCAGCTGAGAAATGATTAAGAAAGCT 5564
723 uargserlyspalaaleuamleuglnserseasnalaaglualeupheg 740
5565 GAGAGCCGCTTGAGAGCTATGAAAAGAGAGCTC...TGCTGTAC 5611
740 lnglnleuasplysleuarg.....Thrlletryaspalaatyral 753
5612 AACACAGAGAGAAAGTACATCATCATCATTTAAGGAGAGAGT 5661
754 lysleuvalglugluthrlleproleualaglulysasnleuasnlnh 770
5662 GAGAACCTTGAAGAGAGCTGAGATGCGAGAGCAAGC...CAGAGCA 5708
770 sleualaspgluserglnlysalaglnalaphaspaspplleugllyv 787
5709 TGCAGGCTTGAGAGCAGAGAAATTCAGAA..... 5736
787 alleualliasvalglnmetaspargaspalavalglualaleuugln 803
5737GAGAGGTAGAGAGCCCTAAAGCAAAATAGAGGAGTACCCAA 5781
804 prothrspthrilleasparghival..... 812
5782 AGCTGAGAGGCTCTGAAATTAAGATGTTGTACTATAAGCTCAGAAAAAGA 5831
813Hlsglulleuglnleuvallysgluvalgluaspleug 826
5832 AAATTCGACAAATGAATTAACAAAAGAGCAAGAGCGAATTCGAAATAG 5881
826 lutyralaleuaspserserlyarggllyal.....lysleu 839
5882 AA...ATAATTAATTCATTCATTTGAATAATTTGCAGAAAGAAAGACAA 5928
840 gluglulleuglnleuasnphelueuglnargtrgshaptryle 856
5929 GAGAAATACAGATG.....AAAGCAAAATCAACAC 5960
856 ullevalgluvalaspasppluarargaspghisargmetleuasnlu 873
5961 TGCATGAGATG.....CTCAACACAAATTAAGAGCTCATAGAGA 6004
873 spmetseralaglnvalargtrphisasnalarggluglu...lys 888
6005 GAGTGGCAGCCCTC.....CATATGACCAAGAAAGCTGTAG 6042
889 vallylaserlelleuglnuargpheuglnlysserlngluginle 905
6043 GCCAAGAGCAGAAAT..... 6057
905 uvalleuallaglnleuglnlysgluginleuallaglnlystysleu 922
6058CTTAGTAGTCAAGTAGAGTGTCTTAAGCTGAGAGGCTCACT 6100
922 eugluginserleuasp...leuserlys 931
6101 TGCACAGAGCCCTTGAGAGGCCCAAAATTAATTAATTTGTTTCATCT 6150
932 glulysgluserleuenglnglutyraasnalaaleuysglulysleuas 948
6151 TCAAGTGAATGCTCATTAAGAGTAGAGATGCAAGAGCAAGAACTGA 6200
948 pgluglutyrlhlsnglnleuallagluarglystargglupheuglnleu 965
6201 GAAGAGAGATGAAGAAATCACTAGACTGAATAATCAATTCAGAACCAAG 6250
965 euaspalaleuugllyargleuasnmetlystleuysglutyryleuasps 981

6251 AGCAGCTTGC..... 6261
 982 LysLysAsnGluLysLysGluLeuGlnGlyArgHisValLeuCysH1 998
 6262TCTAACTGCTCCAGAGTGGAAAGAGACACCACTTTGGAA 6302
 998 sSerGlnLeuGlnSerCysMetalalysGlnArgIleSerAlaGlu 1015
 6303 GGAGCA.....AACTAGAACTGAGAAATCTGCACTGGAAT 6340
 1015 euAsnLysSerLysGluLeuLeuGlnGlyGlnGlnLysArgAsn 1031
 6341 TGGAGCAGAGATCCAGAGTCTCAATCAAA.....AATGCTCT 6381
 1032 IleAspAspAsnLys.....TyrArgLysThrLysAl 1043
 6382 TTCAGACACATTTAGAGTCTGCAGAGCTTCTTCAAGAAATCAAGAA 6431
 1043 aaSPValGlnLeuThrArgAspIleGluSerLeuGlnGluArgLeu 1060
 6432 TGAGCTTGATTCACAAATGACACAAATGCTCTTCTTGGAAAA... 6477
 1060 euSerIleGlySerLeuSerAlaIleGluAlaSPLeuLysArgHisSer 1076
 6478GTAACAAATGACTGCAAGAACTGAGTGCAGAGGAATG 6522
 1077 GlnGlu.....LysGluArgLeuAsnSerG1 1085
 6523 CATGACATGGCAGACAAACAGAGAGCTGCAGAAAGAACTCAGTGAGA 6572
 1085 upheAsnArgTyrGlnGlyThrLeuSerValTyrGlnSerAsnIleSer 1102
 6573 GAAATAGGCTAGCTGAGAGAGTGCAGTACTGCGAAGAAATTAAGA 6622
 1102 ySHisLysGlnGluLeuLysLeuSerGlnTyrLysAspIleGluLysArg 1118
 6623 GCAGCAAAAGATCAATTGAAG..... 6642
 1119 TyrThrAsnGlnPheLeuGlnLeuLysThrThrGluMetAlaAsnLysAs 1135
 6643GAGCTCACACTAGAAATAGTGA..... 6666
 1135 pleuAspArgTyrTyrThrAlaLeuAspLysAlaLeuMetArgPheHis 1152
 6667TTGAAGAGAGCTTAGATTCATGATGACA 6694
 1152 erMetLysMetGlnGluLysAsnLysIleIleLysGluLeuTyrGlnGln 1168
 6695 AAGACCAAGTGAAGAAAGAGAGAGTGAAGAAATAGCTGAATAT 6744
 1169 ThrTyrArgGlyGlnAspIleAsp.....TyrIleSerIleAsnSerAs 1183
 6745 CAGCTACAGCTTCATGAGCTGAAAGAAACACAGCTTTCCTTTGGA 6794
 1183 pSerGlnGlyAlaGlyThrArgSerTyrSerTyrArgValAlaMetGlnT 1200
 6795 CACA.....AACACAGATGAGAGAAATCCAGA 6826
 1200 hGlyAspAlaGluLeuGluMetArgGlyArgCysSerAlaGlyLys 1216
 6827 CATACCGAGAGAAATGACTCTAAAGAAATGCTCAGTTCCAGGAAG 6876
 1217 ValLeuAlaSerLeuIle 1222
 6877 CTGGAGATAGACCTTTTA 6894

seq_name: /cgn2_6/ptodata/2/1na/5A_COMB.seq:us-08-353-700-2

seq_documentation_block:
 ; Sequence 2, Application US/08353700
 ; Patent No. 5599919
 ; GENERAL INFORMATION:
 ; APPLICANT: YEN, TIMOTHY J.

APPLICANT: RATTNER, JEROME B.
 TITLE OF INVENTION: NUCLEIC ACID ENCODING A
 TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,
 TITLE OF INVENTION: AND METHODS OF USE
 NUMBER OF SEQUENCES: 4
 CORRESPONDENCE ADDRESS:
 ADDRESSES: DANN, DOREMAN, HERRELL AND SKILLMAN
 STREET: 1601 MARKET STREET, SUITE 720
 CITY: PHILADELPHIA
 STATE: PA
 COUNTRY: USA
 ZIP: 19103-2307
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/353,700
 FILING DATE: 09-DEC-1994
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: REED, JANET E.
 REGISTRATION NUMBER: 36,252
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (215) 563-4100
 TELEFAX: (215) 563-4044
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 10136 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA to mRNA
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: HUMAN
 US-08-353-700-2

alignment_scores:
 Quality: 439.50 Length: 1256
 Ratio: 0.668 Gaps: 49
 Percent Similarity: 52.389 Percent Identity: 19.904

alignment_block:
 US-09-538-396-2 x US-08-353-700-2 ..
 Align seg 1/1 to: US-08-353-700-2 from: 1 to: 10136

51 CysThrGlyGluLeuProAsnSerArgSerGlyHisThrPheValH1 67
 5619 TGCTCTGGGAGACAGTCCAGATACCAT.....TA 5650
 67 aaSPProLysValAlaGlyGluThrGluThrLysGlnIleLysLeuA 84
 5651 TGAGCT.....CCAGGGAAGATTAACCCAGGAGCTCTTCAGAA... 5690
 64 rghelysThrAlaAlaGlyLysAspValValCysIleArg.....Ser 98
 5691TGCAATTCAGAAATGTC 5708
 99 PheGlnLeuThrGlnLysAlaSerLysMetGluPhe.....Lys 111
 5709 TTTTCTGCTCCATATGCTTGTACTATGATTCCTCGGGAATCAGGA 5758
 111 salalIleGluSerValLeuGlnThrIleAsnProHisThrGlyGluYsV 128
 5739 AGATATCCATTAATCTCACTCGGAGTAAGACATCAATAGAAAT 5808
 128 aLcysLeuSerTyrArgCysAlaAspMetAspArgGluLeuProAlaLeu 144


```

707 GluAspGluPheValLysGlnArgMetGlnAsnSerThrAlaGlu 723
    |||||
738 GAGGAAGAG.....CATCAGCTGAGAAATAGCATTAAGAAAGCT 7375
723 uArgSerLysAlaLeuAlaMetGluSerSerAsnAlaGluAlaLeuPheG 740
    |||||
7376 GAGAGCCCGCTGAGAACCTGATGAAAGAACACCTC...TGTCCTTAC 7422
740 LngLysAspLysLeuArg.....ThrIleTyraAspAlaTyrVal 753
    |||||
7423 AACACCTGAGGAAGAGTGCATCATCATCAGATTACTTAAGGATGAGTGT 7472
754 LysLeuValGluGluThrIleProLeuAlaGluLysAsnLeuSnglnH 770
    |||||
7473 GAGAACCTTGAAAGAGAGCTAGAGATAGCCAGACAAAC...CAAGAGCA 7519
770 sLeuAlaAspGluSerGlnLysAlaGlnAlaPheAspLeuGluLys 787
    |||||
7520 TGCAGCTCTTGAGCGAGAGATTCCAA..... 7547
787 allLeuAlaHisValGlnMetAspArgAspAlaValGluAlaLeuGln 803
    |||||
7548 ....GAGAGCTAGAGACCTTAAGCAAAATAGAGGATGAGCCAA 7592
804 ProThrAspThrIleAspArgHisVal..... 812
7593 AGTCTGAGAGCTCTGGAATTAGATGTGTACTATAAGTCAGAAAAAGA 7642
813 .....HisGluIleGlnGlnLeuValLysGluValGluAspLeuG 826
7643 AAATCTGACAAATGAAATTACAAAGAGCGAAGCGAATCTGAAATTAG 7692
826 LutyraLalaLeuAspSerSerGlyArgGlyVal.....LysSerLeu 839
    |||||
7693 AA...ATAATAATTATCATCTTTGAAATAATTTTGCAAGAAAAAGACAA 7739
840 GluGluIleGlnLeuGluLeuAsnPheLeuGlnArgThrArgAspThr 856
    |||||
7740 GAGAAAGTACAGATG.....AAGCAAAATCAACAC 7771
856 uIleValGluValAspAspLeuArgAspGlnHisArgMetLeuAsnGlu 873
    |||||
7772 TGCATGAGGATG.....CTTCAACACAAATTAAGAGCTCAATGAGA 7815
873 sPheSerSerAlaGlnValArgTrpHisAsnAlaArgGluGlu...Lys 888
7816 GAGTGCACGCCCTG.....CATATGACCAAGAGCCCTTAAG 7853
889 ValLysAlaSerSerIleLeuGlnArgPheGlnLysSerGluGluLe 905
    |||||
7854 GCCAAAGACAGAAAT..... 7868
905 uValLeuLeuAlaGluGluLysGlnLeuIleValGluLysLysLeu 922
    |||||
7869 .....CTTAGTAGCAGTAGAGTGTCTTGAATTGAGAGGCTCAG 7911
922 euGluGluSerLeuAspPro.....LeuSerLys 931
    |||||
7912 TGCCTCAAGGCTTGATGAGGCCAAATAATATATTTGTTGCAATGT 7961
932 GluLysGluSerLeuLeuGlnGluLysAsnAlaLeuLysGlnLysLeu 948
    |||||
7962 TCACGTAAAGGCTCATTTCAAGAGTAGAGATGGCAACAGCAAGCTGGA 8011
948 pGluGluLysHisGlnLeuAlaGluArgLysArgGluPheGlnGlnL 965
    |||||
8012 GAAGAGAGATGAAGAAATCAGTAGACTGAATAATCAATTCAGACCAAG 8061
965 euAspAlaLeuGluLysArgLeuAsnMetLysIleLysGlyTyrLeuAspSer 981
    |||||
8062 AGCAGCTTGTTC..... 8072
982 LysLysAsnGluLysLeuLysGluLeuGlnGluArgHisValLeuGly 998

```

```

8073 .....CTTAACCTGCTCCAGGTGGAAGAGAGACCAACTTTGGA 8113
998 sSerGlnLeuGlnSerCysMetAlaLysGlnAlaArgIleSerAlaGlu 1015
    |||||
8114 GGACCAA.....AACTTAGAAGTGAAGAAATGACGGTGGAT 8151
1015 euAsnLysSerLysGluLeuLeuGlnGlnGlnGlnLysLysArgAsn 1031
    |||||
8152 TGGAGCAGAAATCCAGTCTCAATCCAA.....AATGCCCT 8192
1032 IleAspAsnLeuLys.....TyrArgLysThrLys 1043
    |||||
8193 TTGCAGACACATTTGAAGTCTCAGAGTTCTTACAGAAATCTAGAGAA 8242
1043 AspValGluGlnLeuThrArgAspIleGluSerLeuGluGluArgLeu 1060
    |||||
8243 TGACCTTGAATTTGACAAATAATGCAAAATGCTCTTGTGTAATAA... 8288
1060 euSerIleGlySerLeuSerAlaIleGluAlaAspLeuLysArgHisSer 1076
    |||||
8289 .....GTAAACAAATGACTGCAAGAAAGAACTGAGCTGACAGGAAATG 8333
1077 GlnGlu.....LysGluArgLeuAsnSerG 1085
    |||||
8334 CATAGATGCGCACGAAACAGCAGACCTGCAGAGAACTGCTGAGAGA 8383
1085 uPheAsnArgTrpGlnGlyThrLeuSerValTyrGlnSerAsnIleSer 1102
    |||||
8384 GAAATAATAGCTAGCTGAGAGTTCAGTTACGTTGGAAGAAATAGAA 8433
1102 yHisLysGlnGluLeuLysLeuSerGlnTyrLysAspIleGluLysArg 1118
    |||||
8434 GCAGCAAAAGATCAATTGAA..... 8453
1119 TyrThrAsnGlnPheLeuGlnLeuLysThrThrGluMetAlaAsnLysAs 1135
    |||||
8454 .....GAGCTACACATGAGAAATAGGAA..... 8477
1135 pLeuAspArgTyrThrAlaLeuAspLysAlaLeuMetArgPheHis 1152
    |||||
8478 .....TTGAAGAAAGACCTTAGATGATGATGAC 8505
1152 eMetLysMetGluGluIleAsnLysIleIleLysGluLeuTrpGln 1168
    |||||
8506 AAGACCAAGTGGAAAGAGAAAGTGAAGAGAAATAGCTGAATAT 8555
1169 ThrTyrArgGlyGlnAspIleAsp.....TyrIleSerIleAsnSer 1183
    |||||
8556 CAGCTACGCTTCATGAACTGAAAGAAACACAGGCTTGTGTTTGA 8605
1183 pSerGluGlyAlaGlyThrArgSerTyrSerTyrArgValIleMetGln 1200
    |||||
8606 CAA.....AACAAACAGTTGAGAGTGAAGTGA 8637
1200 hGlyAspAlaGluLeuGluMetArgGlyArgCysSerAlaGlyGln 1216
    |||||
8638 CATACCGAGAGAAATGACTTCAAGAAAGATGTCTCAGTTCACAGAG 8687
1217 ValLeuAlaSerLeuIle 1222
    |||||
8688 CTGAGATGAGACCTTTTA 8705

```

seq_name: /cgn2_6/ptodata/2/ina/PCrus_COMB.seq:PCr-US95-16216-2

seq_documentation_block:

Sequence 2, Application PC/TUS9516216

GENERAL INFORMATION:

APPLICANT: Yen, Timothy J.

APPLICANT: Ratner, Jerome B.

TITLE OF INVENTION: Nucleic Acid Encoding a Transiently

NUMBER OF SEQUENCES: 4 Expressed Kinetochores Protein, and Methods of Use